Tapestry of dualities in decohered quantum error correction codes

Kaixiang Su,[1,*] Zhou Yang [ORCID],[2,*] and Chao-Ming Jian[2]

[1]*Department of Physics, University of California, Santa Barbara, California 93106, USA*
[2]*Department of Physics, Cornell University, Ithaca, New York 14853, USA*

Quantum error correction (QEC) codes protect quantum information from errors due to decoherence. Many of them also serve as prototypical models for exotic topological quantum matters. Investigating the behavior of the QEC codes under decoherence sheds light on not only the codes' robustness against errors but also new out-of-equilibrium quantum phases driven by decoherence. The phase transitions, including the error threshold, of the decohered QEC codes can be probed by the systems' Rényi entropies S_R with different Rényi indices R. In this paper, we study the general construction of the statistical models that characterize the Rényi entropies of QEC codes decohered by Pauli noise. We show that these statistical models can be organized into a "tapestry" woven by rich duality relations among them. For Calderbank-Shor-Steane (CSS) codes with bit-flip and phase-flip errors, we show that each Rényi entropy is captured by a pair of dual statistical models with randomness. For $R = 2, 3, \infty$, there are additional dualities that map between the two error types, relating the critical bit-flip and phase-flip error rates of the decoherence-induced phase transitions in the CSS codes. For CSS codes with an "*em* symmetry" between the X-type and the Z-type stabilizers, the dualities with $R = 2, 3, \infty$ become self-dualities with *super-universal* self-dual error rates. These self-dualities strongly constrain the phase transitions of the code signaled by $S_{R=2,3,\infty}$. For general stabilizer codes decohered by generic Pauli noise, we also construct the statistical models that characterize the systems' entropies and obtain general duality relations between Pauli noise with different error rates.

I. INTRODUCTION

Quantum error correction (QEC) is a fundamental scheme in quantum computation to protect quantum information from decoherence caused by errors and noise [1,2]. A general QEC code achieves this protection by encoding the quantum information into the "logical qubits" formed by many-body states of physical qubits with entanglement patterns resilient against errors. From the quantum matter perspective, a large class of QEC codes, epitomized by the toric code [3], can be viewed as systems with topological orders, where the long-range-entangled and topologically protected ground states serve as the logical qubits. Recently, there has been tremendous progress in the experimental realizations of the toric code and related QEC codes on noisy intermediate-scale quantum platforms [4–7]. The preparation of a QEC code on a noisy quantum platform is, in general, an out-of-equilibrium process inevitably subject to decoherence. Instead of a Gibbs ensemble, the prepared state is an "error-corrupted" mixed state, i.e., a classical mixture of different error patterns on top of the code's logical states. Understanding such error-corrupted mixed states brings insights into the robustness of QEC codes against errors and the effect of decoherence on topological orders.

An important metric for the performance of a QEC code is the error threshold, which pertains to the "decodability" of the logical qubits from the error-corrupted mixed state. The seminal work Ref. [8] pointed out that the decodability of the 2D toric code with errors is captured by the 2D classical random-bond Ising model, and the error threshold is identified as a continuous phase transition. This method that maps the QEC code's decodability to classical statistical models has been widely generalized to other stabilizer codes (see Refs. [9–13] for examples).

Recent works Refs. [14–16] provide a different perspective on the error threshold of the toric code by viewing it as a singularity intrinsic to the error-corrupted mixed state ρ. This singularity manifests a decoherence-induced phase transition (DIPT) of the system's total von Neumann entropy $S_1 = -\mathrm{Tr}\rho \log \rho$ as a function of the decoherence strength, i.e., the error rate. Moreover, the error threshold naturally belongs to the rich family of DIPTs of the Rényi entropies $S_R = \frac{1}{1-R} \log(\mathrm{Tr}\rho^R)$ (for $R = 2, 3, ...$ and $R \to 1$) in the decohered toric code [14–16]. For a general QEC code under decoherence, the DIPTs of the Rényi entropies probe the singular changes in the entanglement pattern in the error-corrupted mixed state. In light of the close relation between QEC codes and topological orders in systems free of decoherence, studying the behavior of the error-corrupted mixed state in decohered QECs and the DIPTs therein provides an interesting path toward understanding topological orders and other decoherence-induced exotic phases in beyond-Gibbs-ensemble mixed states.

In this paper, we focus on developing a general framework to study the error-corrupted mixed states in stabilizer codes

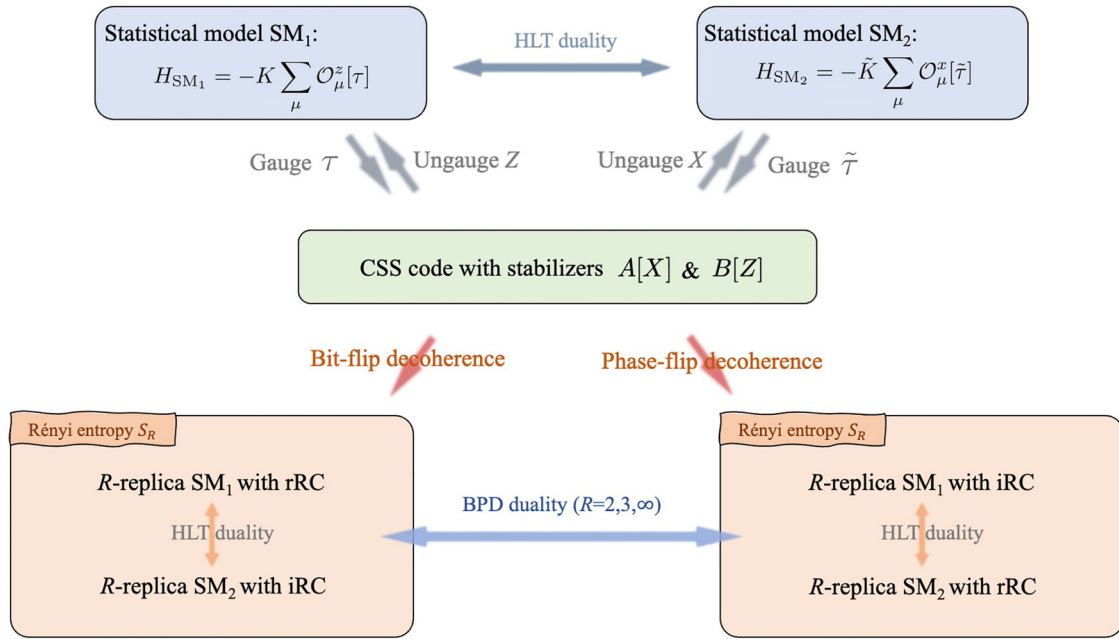*These authors contributed equally to this work.

FIG. 1. The tapestry of dualities of a general CSS code. In Sec. II B, we introduce the statistical models $SM_{1,2}$ through ungauging the CSS code and introduce the R-replica $SM_{1,2}$ with randomness (rRC and iRC) as the descriptions of the Rényi entropies of the decohered CSS code. The HLT and BPD dualities are discussed in Secs. II B and II C.

decohered by Pauli noise, including bit-flip and phase-flip errors in particular. We systematically construct the statistical models arising from the Rényi entropies S_R for a general R and identify a rich set of duality relations that weave these statistical models into a "tapestry of dualities." Our discussion contains three major parts, focusing on three classes of translation-invariant stabilizer codes: (1) general Calderbank-Shor-Steane (CSS) codes, (2) CSS codes with an "*em* symmetry" between the X-type and the Z-type stabilizers, and (3) general stabilizer codes.

We first investigate general CSS codes decohered by bit-flip and phase-flip errors. The tapestry of dualities for this class of decohered codes is shown in Fig. 1. The two error types cause independent decoherence in a CSS code and are, hence, studied separately. The behavior of the decohered CSS code is intimately related to a pair of dual statistical models, named SM_1 and SM_2, which are obtained from ungauging the CSS code. Under bit-flip errors, the Rényi entropy S_R of the error-corrupted mixed state is described by a pair of dual statistical models with randomness: (1) R-replica SM_1 with real random couplings (rRC) and (2) R-replica SM_2 with imaginary random couplings (iRC). These two statistical models are dual under a high-low-temperature (HLT) duality. Similarly, under phase-flip errors, S_R is described by another pair of random statistical models: (1) R-replica SM_1 with iRC and (2) R-replica SM_2 with rRC, which are again HLT dual to each other. The patterns of randomness in these statistical models are interchanged as we switch from bit-flip errors to phase-flip errors. Additionally, for $R = 2, 3, \infty$, we discover extra duality relations that map between (the random statistical models associated with) bit-flip and phase-flip errors. We call these dualities the bit-phase-decoherence (BPD) dualities. The random statistical models and the dualities provide powerful tools to study the phases and the DIPTs in the decohered

CSS codes. We propose a conjecture on the monotonicity of the DIPTs' critical error rates as functions of the Rényi index R, and discuss the alternative interpretation of the DIPT with $R = 2$ as a quantum phase transition in the doubled Hilbert space. We also discuss two concrete examples, the decohered 3D toric code [17] and the decohered X-cube model [18], demonstrating the abovementioned general results.

When the CSS codes are endowed with additional symmetries, the duality structure under decoherence can be enriched. We analyze the case when the CSS codes have a symmetry that exchanges the X-type and Z-type stabilizers, which we dub "electric-magnetic symmetry" (*em* symmetry). Under the bit-flip and phase-flip errors, the tapestry of dualities of the *em*-symmetric CSS code is shown in Fig. 7, which is effectively the tapestry shown in Fig. 1 "folded in half" by the symmetry. The *em* symmetry interchanges the bit-flip and phase-flip errors. The original pair of dual statistical models SM_1 and SM_2 become the same model, call it SM, with a self-duality. Under either bit-flip or phase-flip errors, the Rényi entropies S_R are captured by a pair of statistical models with randomness: (1) R-replica SM with rRC and (2) R-replica SM with iRC. They are related by the HLT duality for any R. The BPD dualities for $R = 2, 3, \infty$ (combined with the *em* symmetry) become self-dualities with *super-universal* self-dual error rates shared by all *em*-symmetric CSS codes in different dimensions. For a given *em*-symmetric CSS code (and a fixed $R = 2, 3,$ or ∞), if there is a unique DIPT of S_R as the error rate varies, then the critical error rate must coincide with the super-universal self-dual values. We demonstrate our general results in the concrete settings of the decohered 2D toric code [17], and the decohered Haah's code in 3D [19].

For general stabilizer codes, bit-flip and phase-flip errors are no longer special. We therefore consider the decoherence by general Pauli noise. The general decohered stabilizer codes

are tied to a single statistical model SM, which is self-dual under an HLT duality. The Rényi entropy S_R of the decohered code is captured by both (1) the R-replica SM with rRC and (2) the R-replica SM with iRC, two random statistical models related by an HLT duality. Moreover, we find extra dualities between different error rates for $R = 2, \infty$ and call them "general-Pauli-noise dualities" (GPN dualities). The GPN dualities are the generalizations of the BPD dualities of decohered CSS codes. We obtain the surface of super-universal self-dual error rates under the GPN dualities.

The rest of the paper is organized as follows. Section II focuses on the general CSS codes decohered by bit-flip and phase-flip errors. Section III discusses the decohered CSS codes with *em* symmetry. Section IV analyzes general stabilizer codes with general Pauli noise. We present our conclusion and outlook in Sec. V.

II. CSS CODE UNDER DECOHERENCE

In this section, we will first review the basics of general CSS codes and introduce the decoherence model that describes bit-flip and phase-flip errors. Next, we provide the general construction of the statistical models, the R-replica SM_1 and SM_2 with randomness, that describe the Rényi entropies S_R of the error-corrupted mixed states of the decohered CSS code. We will address how the $R \to 1$ limit recovers the disordered statistical model that captures the code's error threshold for decodability. Then, we present the HLT and BPD dualities between these statistical models and the tapestry of dualities they form. The DIPTs are signaled by the singular behavior of the Rényi entropies S_R and, hence, are identified with the phase transitions in the random statistical models. For $R = 2$, by generalizing Refs. [15,20], we give an alternative interpretation of the DIPT as a quantum phase transition in the doubled Hilbert space. We also propose a conjecture on the monotonicity of the DIPTs' critical error rates as a function of the Rényi index R. Finally, we study the decohered 3D toric code and the decohered X-cube model as concrete examples.

A. Introduction to CSS codes and error models for decoherence

First, we briefly review the general description of a CSS code. Consider a set of N qubits (labeled by μ) arranged into a lattice. A general CSS code is a stabilizer code with two types of stabilizers: X-type and Z-type. Each X-type generator, denoted as $A_i[X]$, is a product of only Pauli-X operators X_μ, while each Z-type generator, denoted as $B_j[Z]$, is a product of only Pauli-Z operators Z_μ. All the stabilizers commute with each other:

$$[A_i[X], B_j[Z]] = 0, \quad \forall i, j. \tag{1}$$

The logical subspace of the CSS code is the ground-state Hilbert space of the CSS code Hamiltonian

$$H_{\text{css}} = -\sum_i A_i[X] - \sum_j B_j[Z]. \tag{2}$$

Each logical state $|\Omega\rangle$ satisfies

$$A_i[X] |\Omega\rangle = B_j[Z] |\Omega\rangle = |\Omega\rangle, \quad \forall i, j. \tag{3}$$

For this work, we work with CSS codes with the following general properties. We assume that H_{css} is translation invariant (modulo boundary effects of the lattice if they exist) and all the terms in H_{css} are local. The number of $A_i[X]$ terms does not have to equal the number of $B_j[Z]$ terms. But we assume that there is no local logical operator and the rate of the code N_c/N vanishes in the large system limit. Here, N_c denotes the total number of logical qubits. These properties are rather general. They are broadly satisfied by commonly discussed topological CSS codes, including the toric code [17], the surface code [8,21], color codes [22], the X-cube model [18], Haah's code [19], etc.

Next, we introduce the error models for the decoherence that we will study in this section. We consider an error model with bit-flip errors and phase-flip errors. For a single site μ, when a bit-flip (phase-flip) error occurs, the system's state is acted on by X_μ (Z_μ). The classical probabilities p_x and p_z for the appearance of these errors on a single qubit are called the error rates. The decoherence caused by these two types of errors is described by the following quantum channels, respectively:

$$\mathcal{N}_{x,\mu}(\rho) = (1 - p_x)\rho + p_x X_\mu \rho X_\mu,$$
$$\mathcal{N}_{z,\mu}(\rho) = (1 - p_z)\rho + p_z Z_\mu \rho Z_\mu. \tag{4}$$

When all the physical qubits are subject to such decoherence, a logical state $\rho_0 = |\Omega\rangle\langle\Omega|$, initially pure, becomes an error-corrupted mixed-state:

$$\rho_0 \to \mathcal{N}_x \circ \mathcal{N}_z(\rho_0), \tag{5}$$

a mixture of the logical state dressed by all error patterns. Here, we have defined the total decoherence channels $\mathcal{N}_{x/z} \equiv \otimes_\mu \mathcal{N}_{x/z,\mu}$. It suffices to consider the range of error rates $0 \leqslant p_x, p_z \leqslant \frac{1}{2}$ because the system experiences the maximal amount of bit-flip (phase-flip) decoherence when $p_x = 1/2$ ($p_z = 1/2$).

The bit-flip errors only affect the Z-type stabilizers, while the phase-flip errors only influence the X-type stabilizers. Hence, the two error types cause independent decoherence effects (more details in the next subsection). Without loss of generality, it suffices to independently study the two types of mixed states ρ_b and ρ_p corrupted only by the bit-flip and phase-flip errors, respectively:

$$\rho_b = \mathcal{N}_x(\rho_0), \quad \rho_p = \mathcal{N}_z(\rho_0). \tag{6}$$

We can write these mixed states as a summation over error chains. For example,

$$\rho_b = \sum_E P_x(E) X_E \rho_0 X_E, \tag{7}$$

where $\sum_E$ is a summation over all possible error chain E. An error chain E is the set of qubits where the errors occur and $X_E \equiv \prod_{\mu \in E} X_\mu$. $P_x(E)$ is the probability for the error chain E to appear:

$$P_x(E) = p_x^{|E|}(1 - p_x)^{N-|E|}. \tag{8}$$

Here, $|E|$ denotes the number of qubits in E. Similarly, we can write

$$\rho_p = \sum_E P_z(E) Z_E \rho_0 Z_E, \tag{9}$$

with $P_z(E) = p_z^{|E|}(1 - p_z)^{N-|E|}$ and $Z_E \equiv \prod_{\mu \in E} Z_\mu$. In the next subsection, we develop the statistical models that capture the behavior of the error-corrupted mixed states $\rho_{\mathrm{p/b}}$.

As a preparation for the subsequent analysis, it is helpful to introduce an N-dimensional $\mathbb{Z}_2$ vector space $\mathcal{V}$ for the lattice of N qubits. Any subset W of qubits can be represented as an N-component vector $W \in \mathcal{V}$ whose μth entry is 1 if the μth site belongs to the subset W and 0 if not. Here and for the rest of the paper, we use the same notation for a subset of qubits and its corresponding vector. For a subset W, its cardinality $|W|$ counts the number of nonzero entries in the vector W. With two vectors $W_1, W_2 \in \mathcal{V}$, $W_1 + W_2$, and $W_1 \cdot W_2$ represent the addition and the dot product in the $\mathbb{Z}_2$ vector space $\mathcal{V}$. In terms of subsets, $W_1 + W_2$ represents $W_1 \cup W_2 - W_1 \cap W_2$ and $W_1 \cdot W_2$ is the parity of $|W_1 \cap W_2|$.

For each stabilizer $A_i[X]$, we denote the subset of qubits it includes as a_i. Similarly, every $B_j[Z]$ has a corresponding vector b_j. The commutation relation Eq. (1) is equivalent to

$$\mathsf{a}_i \cdot \mathsf{b}_j = 0 \quad \forall i, j. \tag{10}$$

The collection of vectors $\{\mathsf{a}_i\}$ ($\{\mathsf{b}_j\}$) span a sub vector space denoted as $\mathcal{V}_x$ ($\mathcal{V}_z$). The relation Eq. (10) implies that $\mathcal{V}_x$ belongs to the orthogonal subspace of $\mathcal{V}_z$, namely $\mathcal{V}_x \subset \mathcal{V}_z^\perp$. Similarly, we have $\mathcal{V}_z \subset \mathcal{V}_x^\perp$. A logical X operator (Z operator) of the CSS code must be a product of X_μ (Z_μ) on a subset of qubits whose corresponding vector belongs to the quotient space $\mathcal{V}_z^\perp/\mathcal{V}_x$ ($\mathcal{V}_x^\perp/\mathcal{V}_z$). The number of logical qubits is then given by $N_\mathrm{c} = \dim(\mathcal{V}_z^\perp/\mathcal{V}_x) = \dim(\mathcal{V}_x^\perp/\mathcal{V}_z)$. Our assumption that the code rate $N_c/N \to 0$ in the infinite system limit conceptually implies $\mathcal{V}_z^\perp$ ($\mathcal{V}_x^\perp$) is almost the same as $\mathcal{V}_x$ ($\mathcal{V}_z$), which is important for the dualities we will discuss later. The assumption that the CSS code is free of local logical operators implies that the quotient spaces $\mathcal{V}_z^\perp/\mathcal{V}_x$ and $\mathcal{V}_x^\perp/\mathcal{V}_z$ only contain vectors that represent nonlocal sets that involve infinitely many qubits in the infinite system limit.

In addition to characterizing the stabilizers, it proves convenient to treat any error chain E as a vector in $\mathcal{V}$ when analyzing the effects of decoherence. For example, the error chains E, E' follow the multiplication and commutation relations,

$$X_E X_{E'} = X_{E+E'}, \quad Z_E Z_{E'} = Z_{E+E'},$$
$$X_{E'} Z_E = (-1)^{E \cdot E'} Z_E X_{E'}. \tag{11}$$

For any subset E, if $E \in \mathcal{V}_x$, then X_E can be written as a product of the $A_i[X]$ stabilizers. Similarly, Z_E is product of the $B_j[Z]$ stabilizers if and only if $E \in \mathcal{V}_z$. The summation $\sum_E$ in Eqs. (7) and (9) can be viewed as a summation $\sum_{E \in \mathcal{V}}$ over all $\mathbb{Z}_2$ vectors in $\mathcal{V}$.

B. Statistical models for decohered CSS codes

We now develop the systematic framework to characterize the effects of decoherence in the error-corrupted mixed state $\rho = \mathcal{N}_x \circ \mathcal{N}_z(\rho_0)$ using statistical models. These decoherence effects are manifested in the behavior of $\mathrm{Tr}\rho^R$, which are directly related to the Rth Rényi entropies $S_R = \frac{1}{1-R} \log(\mathrm{Tr}\rho^R)$. Singularities in S_R (as functions of the error rates p_x and p_z) signal the DIPTs. In particular, the singularity in the limit $R \to 1$ is the error threshold of the decodability of the code.

In the following, we show that $\mathrm{Tr}\rho^R$ can be formulated as the partition functions of a family of statistical models of classical $\mathbb{Z}_2$ spins with random couplings. The Rényi entropies are, therefore, the respective free energies. The DIPTs can then be identified as the phase transitions in these statistical models. This subsection is devoted to the general construction of these statistical models for the decohered CSS code in the infinite-system limit. Concrete examples will be provided in the subsequent subsections Secs. II E and II F.

For a CSS code $\mathcal{C}$, decoherence effects of the bit-flip errors decouple from those of the phase-flip errors. The reason is that the former only affects the $B_j[Z]$ stabilizers while the latter only affects the $A_i[X]$ stabilizers. The decoupling is manifested by the following identity in the infinite-system limit

$$\mathrm{Tr}\rho^R = \mathrm{Tr}\rho_\mathrm{b}^R \times \mathrm{Tr}\rho_\mathrm{p}^R. \tag{12}$$

In infinite-system limit, the absence of local logic operators implies $|\langle\Omega| X_E Z_{E'} |\Omega\rangle| = |\langle\Omega| X_E |\Omega\rangle| \times |\langle\Omega| Z_{E'} |\Omega\rangle|$ for the logical state $|\Omega\rangle$ and any pair of finite subsets of qubits E, E'. This property leads to Eq. (12). The decoupling of the bit-flip and phase-flip errors allows us to independently formulate the statistical models for $\mathrm{Tr}\rho_\mathrm{b}^R$ and $\mathrm{Tr}\rho_\mathrm{p}^R$ without loss of generality. Also, note that the absence of local logical operators makes all logical states locally indistinguishable. For this reason, different choices of the error-free logical state $\rho_0 = |\Omega\rangle \langle\Omega|$ will not affect the construction of the statistical models in the infinite-system limit.

As the first step toward the statistical models, we use Eqs. (7) and (9) to expand $\mathrm{Tr}\rho_\mathrm{b/p}^R$ as a sum over the vector spaces $\mathcal{V}, \mathcal{V}_x$, and $\mathcal{V}_z$. Take $\mathrm{Tr}\rho_\mathrm{b}^2$ as an example. In the infinite-system limit,

$$\mathrm{Tr}(\rho_\mathrm{b}^2) = \sum_{E,E'\in\mathcal{V}} P_x(E)P_x(E')|\langle X_E X_{E'}\rangle_\Omega|^2$$
$$= \sum_{E\in\mathcal{V}} \sum_{C\in\mathcal{V}_x} P_x(E)P_x(E + C). \tag{13}$$

Here, we have used the fact that $|\langle X_E X_{E'}\rangle_\Omega|^2 = 0$ unless $X_E X_{E'} = X_{E+E'}$ can be written as a product of the $A_i[X]$ stabilizers, i.e., $E + E' \in \mathcal{V}_x$. When $E + E' \in \mathcal{V}_x$, we have $|\langle X_E X_{E'}\rangle_\Omega|^2 = 1$ and $E' = E + C$ for some $C \in \mathcal{V}_x$. For general R, we have

$$\mathrm{Tr}(\rho_\mathrm{b}^R) = \sum_{E\in\mathcal{V}} \sum_{C_2,\ldots R\in\mathcal{V}_x} P_x(E)P_x(E + C_2)\ldots P_x(E + C_R)$$
$$= \frac{1}{2^{\dim\mathcal{V}_x}} \sum_{E\in\mathcal{V}} \left(\sum_{C_{1,2,\ldots,R}\in\mathcal{V}_x} P_x(E + C_1)\ldots P_x(E + C_R) \right). \tag{14}$$

The last line is obtained from the second by shifting $E \to E + C_1, C_{2,\ldots,R} \to -C_1 + C_{2,\ldots,R}$ with $C_1 \in \mathcal{V}_x$.

Later, we will see that $\mathrm{Tr}(\rho_\mathrm{b}^R)$ can be mapped to the partition function of an R-replica statistical model of classical $\mathbb{Z}_2$ spins with random couplings. $E \in \mathcal{V}$ specifies a pattern of the random couplings, and the partition function of each replica is given by $\sum_{C\in\mathcal{V}_x} P_x(E + C)$ (before averaging/summing over all randomness patterns E). The interaction between different replicas is only mediated by the random coupling E.

In the $R \to 1$ limit, the Rényi entropy S_R reduces to the von Neumann entropy S_1:

$$S_1(\rho_\mathrm{b}) = \lim_{R\to 1} \frac{1}{1-R} \big(\mathrm{Tr}\big(\rho_\mathrm{b}^R\big) - 1 \big)$$

$$= -\sum_{E\in\mathcal{V}} P_x(E) \log\left(\sum_{C\in\mathcal{V}_x} P_x(E+C)\right), \qquad (15)$$

which turns out to be the quenched-disorder-average of the free energy $-\log(\sum_{C\in\mathcal{V}_x} P_x(E+C))$ in the random statistical model.

Similar to $\mathrm{Tr}(\rho_\mathrm{b}^R)$, we can expand $\mathrm{Tr}(\rho_\mathrm{b}^R)$ as

$$\mathrm{Tr}\big(\rho_\mathrm{p}^R\big) = \frac{1}{2^{\dim\mathcal{V}_z}} \sum_{E\in\mathcal{V}} \left(\sum_{C_{1,2,\dots,R}\in\mathcal{V}_z} P_z(E+C_1)\dots P_z(E+C_R)\right),$$
$$(16)$$

which will also be mapped to the partition function of certain R-replica statistical models with random couplings. Consequently, the von Neumann entropy $S_1(\rho_\mathrm{p}) = \lim_{R\to 1} S_R(\rho_\mathrm{p})$ will be identified as the quenched-disorder-average free energy of the corresponding statistical model.

In the following, we present the systematic construction of the random statistical models of classical $\mathbb{Z}_2$ spins whose partition functions map to $\mathrm{Tr}(\rho_{\mathrm{b/p}}^R)$. For the sake of clarity, we introduce these statistical models in two steps. First, we construct these statistical models without randomness. These nonrandom statistical models naturally arise from ungauging the CSS code. In the second step, we establish the mapping between the multireplica version of these statistical models (with randomness) and the quantities $\mathrm{Tr}(\rho_{\mathrm{b/p}}^R)$ [which are related to the Rényi entropies $S_R(\rho_{\mathrm{b/p}})$].

Now, we perform the first step of our construction. For a general CSS code $\mathcal{C}$, we introduce two (nonrandom) classical statistical models, denoted as SM_1 and SM_2. They can be obtained naturally by treating the CSS code as a generalized $\mathbb{Z}_2$ gauge theory and then ungauging it. As we will see, this ungauging procedure is exactly the inverse of the gauging procedure introduced in Ref. [18], which constructs CSS codes from statistical models of classical $\mathbb{Z}_2$ spins. In this step, we explain the ungauging as a procedure introduced by hand. When we consider $\mathrm{Tr}(\rho_{\mathrm{b/p}}^R)$ later in the second step, the ungauging is effectively implemented by the errors.

The ungauging procedure follows from the intuition based on the 2D toric code (see Fig. 2), which is equivalent to a conventional $\mathbb{Z}_2$ gauge theory. The general idea is as follows. For a general CSS code $\mathcal{C}$, we can treat the Z_μ's as the gauge field variables. The $A_i[X]$ terms in the CSS code Hamiltonian Eq. (2) implement the generalized Gauss law while the $B_j[Z]$ terms describe the energy costs of gauge fluxes. Ungauging the CSS code $\mathcal{C}$ from this perspective produces the classical statistical model SM_1. If we instead treat X_μ as the gauge field and interchange the roles of $A_i[X]$ and $B_j[Z]$, ungauging the CSS code $\mathcal{C}$ results in classical statistical model SM_2.

We now explicitly carry out the ungauging procedure to construct the nonrandom version of SM_1 from the general CSS code $\mathcal{C}$. We introduce a $\mathbb{Z}_2$ spin variable τ_i at the center of each $A_i[X]$ term and treat it as the matter field coupled to the generalized $\mathbb{Z}_2$ gauge field Z_μ. The operator $A_i[X]\tau_i^x$ for each
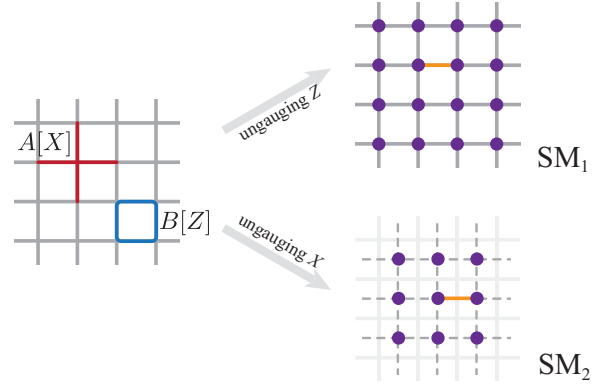

FIG. 2. For the 2D toric code, the two types of stabilizers $A[X]$ and $B[Z]$ are pictorially represented above. Ungauging the gauge field Z and the gauge field ungauging X in the 2D toric code produce the 2D Ising model on the original lattice and that on the dual lattice, respectively. The original square lattice is depicted in gray solid lines, and the dual lattice is depicted in gray dashed lines. Each purple dot represents a classical $\mathbb{Z}_2$ spin $\tau/\tilde{\tau}$, and the orange edge presents the near-neighbor interactions in the Ising models. (see Sec. III B for the microscopic details)

i generates the gauge transformation, where τ_i^x is a Pauli-X operator. The gauge field Z_μ couples to the matter fields via the gauge-invariant interaction $Z_\mu \mathcal{O}_\mu^z[\tau^z]$ with

$$\mathcal{O}_\mu^z[\tau^z] = \prod_i \big(\tau_i^z\big)^{(\mathsf{a}_i)_\mu}. \qquad (17)$$

Recall that $\mathsf{a}_i \in \mathcal{V}$ is the $\mathbb{Z}_2$ vector associated with the stabilizer $A_i[X]$. $(\mathsf{a}_i)_\mu$ is the μth component of a_i. Due to the locality of each stabilizer $A_i[X]$, $\mathcal{O}_\mu^z[\tau^z]$ must be a local term as well for each μ. The commutation relation $[A_i[X]\tau_i^x, Z_\mu \mathcal{O}_\mu^z[\tau^z]] = 0$ for any i, μ implies the gauge-invariance of the interaction $Z_\mu \mathcal{O}_\mu^z[\tau^z]$. Hence, we can write the following Hamiltonian which describes the generalized $\mathbb{Z}_2$ gauge theory coupled to matter fields:

$$H_{\mathrm{gt1}} = -\sum_i A_i[X]\tau_i^x - K \sum_\mu Z_\mu \mathcal{O}_\mu^z[\tau^z] - \sum_j B_j[Z], \quad (18)$$

where $K > 0$ is a coupling constant and the first term $\sum_i A_i[X]\tau_i^x$ leads to an emergent Gauss law (with matter fields) $A_i[X]\tau_i^x = 1$, $\forall i$ at low energies.

If we add an extra term $-g\sum_i \tau_i^x$ with a large positive g to the Hamiltonian H_{gt1}, the matter field τ_i will be gapped out and the CSS code Hamiltonian Eq. (2) will be recovered as the low-energy theory.

However, the model H_{gt1} allows us to ungauge by turning off all the gauge fields, namely setting $Z_\mu = 1$, and removing the Gauss-law terms $-\sum_i A_i[X]\tau_i^x$. The resulting Hamiltonian is the Hamiltonian of SM_1 (up to a constant):

$$H_{\mathrm{SM}_1} = -K \sum_\mu \mathcal{O}_\mu^z[\tau]. \qquad (19)$$

Here, we have suppressed the superscript z for the spin variables τ_i to emphasize that H_{SM_1} is essentially a *classical* Hamiltonian with commuting spin variables $\tau_i = \pm 1$. It is helpful to note that the lattice of the spin variable τ_i of SM_1 differs from the lattice of the qubits labeled by μ.

The construction of SM_2 is parallel to SM_1. We simply need to interchange the roles of the $A_i[X]$ and $B_j[Z]$ stabilizers. For SM_2, a $\mathbb{Z}_2$ spin variable $\tilde{\tau}_j$ is introduced at the center of each $B_j[Z]$ stabilizer. With X_μ treated as the generalized $\mathbb{Z}_2$ gauge field, ungauging the CSS code produces the classical Hamiltonian of SM_2:

$$H_{SM_2} = -\tilde{K} \sum_\mu \mathcal{O}_\mu^x[\tilde{\tau}], \tag{20}$$

where $\tilde{K}$ is the coupling constant of SM_2 and the spin interaction $\mathcal{O}_\mu^x[\tilde{\tau}]$ is given by

$$\mathcal{O}_\mu^x[\tilde{\tau}] = \prod_j (\tilde{\tau}_j)^{(b_j)_\mu}. \tag{21}$$

We emphasize that each $\mathcal{O}_\mu^x[\tilde{\tau}]$ term is a local spin interaction due to the locality of the $B_j[Z]$ stabilizers.

The two (nonrandom) statistical models SM_1 and SM_2 are dual to each other under a Kramers-Wannier-like HLT duality. We defer the details of this duality to the next subsection, where we include this duality as a component of the tapestry of dualities.

With $SM_{1,2}$ introduced, we are ready to establish the mapping between $\text{Tr}\rho_{b/p}^R$ for error-corrupted mixed states $\rho_{b/p}$ and the R-replica $SM_{1,2}$ with random couplings. There are two types of random couplings: real random couplings and imaginary random couplings, which we refer to as rRC and iRC for short.

For the error-corrupted mixed state ρ_b, we obtain the following theorem.

Theorem 1. For the CSS code in the infinite system limit, $\text{Tr}\rho_b^R$ of the mixed state ρ_b with bit-flip errors is proportional to the partition function of the R-replica SM_1 with real random couplings (rRC). It is also proportional to the partition function of the R-replica SM_2 with imaginary random couplings (iRC):

$$\text{Tr}(\rho_b^R) \propto \sum_{E \in \mathcal{V}} (Z_{SM_1}(K, E))^R \propto \sum_{E \in \mathcal{V}} (W_{SM_2}(\tilde{K}, E))^R, \tag{22}$$

where the coupling constants K and $\tilde{K}$ are given by the bit-flip error rate $p_x \in [0, 1/2]$ via $e^{-2K} = \tanh \tilde{K} = \frac{p_x}{1-p_x}$. The proportionality constants only depend on p_x smoothly and, hence, are unimportant for potential phases and DIPTs.

On the first line of Eq. (22), $Z_{SM_1}(K, E)$ is the single-replica partition function of SM_1 in the presence of the rRC configuration labeled by the error chain E:

$$Z_{SM_1}(K, E) = \sum_{\{\tau_i = \pm 1\}} \exp \left(K \sum_\mu (-1)^{E_\mu} \mathcal{O}_\mu^z[\tau] \right). \tag{23}$$

Here, E_μ is the μth component of the $\mathbb{Z}_2$ vector $E \in \mathcal{V}$. E adds extra sign randomness to the real coupling constant K. Hence, the name rRC. Note that $K > 0$ since $p_x \in [0, 1/2]$. One can show that

$$Z_{SM_1}(K, E) = [p_x(1-p_x)]^{-N/2} N_s \sum_{C \in \mathcal{V}_x} P_x(E + C), \tag{24}$$

where N_s is a factor related to internal symmetries of SM_1. We refer to Appendix A for the derivation of this expression. The first line of Eq. (22) is an immediate consequence of

Eqs. (24) and (14). Equation (24) also shows that $Z_{SM_1}(K, E)$ is proportional to the total probability of all the error chains within the same equivalence class $[E]_x \in \mathcal{V}/\mathcal{V}_x$. Physically, the error chains within the same class $[E]_x$ differ from each other only by elements in $\mathcal{V}_x$ and, hence, cause excitations on the same set of $B_j[Z]$ stabilizers.

To provide some heuristics, we argue that the bit-flip errors suppress the coherent quantum fluctuations of the generalized $\mathbb{Z}_2$ gauge field Z_μ. Hence, in the CSS code with bit-flip decoherence, a modified version of Eq. (18) should appear. The modification includes turning off the first term (that generates coherent quantum fluctuations of the gauge field Z_μ) and treating gauge field Z_μ as a classical degree of freedom with values ± 1. After this modification, SM_1 with rRC as defined in Eq. (23) emerges.

On the second line of Eq. (22), $W_{SM_2}(\tilde{K}, E)$ is the single-replica partition function of SM_2 in the presence of the iRC configuration labeled E:

$$W_{SM_2}(\tilde{K}, E)$$
$$= \sum_{\{\tilde{\tau}_j = \pm 1\}} \exp \left(-H_{SM_2} \right) \prod_\mu \left(\mathcal{O}_\mu^x[\tilde{\tau}] \right)^{E_\mu}$$
$$= \sum_{\{\tilde{\tau}_j = \pm 1\}} \exp \left(-H_{SM_2} - \frac{i\pi}{2} \sum_\mu E_\mu (\mathcal{O}_\mu^x[\tilde{\tau}] - 1) \right), \tag{25}$$

where H_{SM_2} is the same as in Eq. (20). For $p_x \in [0, 1/2]$, $\tilde{K}$ is positive. In $W_{SM_2}(\tilde{K}, E)$, a nontrivial E introduces extra imaginary coupling constants for the spin interaction $\mathcal{O}_\mu^x[\tilde{\tau}]$. And hence the name iRC.

Based on Theorem 1, the phases and the DIPTs of the decohered CSS code probed by the Rényi entropy $S_R(\rho_b)$ can be investigated via studying the phases and transitions in the R-replica SM_1 with rRC and in the R-replica SM_2 with iRC. The fact that two different statistical models describe the behavior of the same quantity $\text{Tr}\rho_b^R$ [or $S_R(\rho_b)$] indicates a duality between them. This duality is, in fact, an HLT duality, which we will discuss more about in the next subsection. For the proof of Theorem 1 (including the HLT duality), we refer to Appendix A. The applications of Theorem 1 in concrete examples are provided in Secs. II E and II F.

Using Eq. (24) and taking the $R \to 1$ limit of Theorem 1, we obtain

Corollary 1. The von Neumann entropy of the error-corrupted state ρ_b is given by the quenched-disorder-averaged free energy of SM_1 with rRC (up to an unimportant additive constant):

$$S_1(\rho_b) = -\sum_{E \in \mathcal{V}} P_x(E) \log \left(Z_{SM_1}(K, E) \right), \tag{26}$$

where K is given by the bit-flip error rate p_x via $e^{-2K} = \frac{p_x}{1-p_x}$.

Here, we treat the rRC configuration E as a disorder following the probability distribution $P_x(E)$. Therefore, the right-hand side of Eq. (26) is interpreted as a quenched-disorder-averaged free energy. In a general disordered spin model, the probability distribution of disorder and the coupling constant K can be independent parameters. The relation $e^{-2K} = \frac{p_x}{1-p_x}$ that appears in SM_1 with rRC is known as the Nishimori condition [23,24].

We remark that the right-hand side of Eq. (26) is exactly the (averaged) free energy of the disordered statistical model introduced by Ref. [8] (and later generalized by Ref. [13]) to study the decodability of $\mathbb{Z}_2$ stabilizer codes with Pauli errors (bit-flip errors in this case). In particular, Ref. [8] established that the phase transition (where the quenched-disorder-averaged free energy becomes singular) in the disordered statistical model signals the decoding error threshold for the code. From the perspective of our current work, the error threshold is viewed as the $R \to 1$ limit of the family of DIPTs labeled by R. Recall that the DIPT with index R is associated with the singularity of the Rényi entropy S_R of the decohered code. We generally expect the critical error rate $p_x^\star(R)$ of the DIPT to be a function of R. The bit-flip error threshold for decoding is given by the limit $p_x^\star(R \to 1)$. As a clarification, for a generic R, we use the term "critical error rate" for the error rate where the Rényi entropy S_R develops singularity and the DIPT with index R occurs. The term "(decoding) error threshold" is only associated with the DIPT in the limit $R \to 1$ and is hence given by the critical error rate in the same limit. The same terminology convention applies to all types of errors considered in this paper. We will discuss the physical meaning of DIPTs at different R's and the dependence of the critical error rate $p_x^\star(R)$ on R in Sec. II D.

For the mixed state ρ_p corrupted by phase-flip errors, we prove a similar theorem that maps $\text{Tr}\rho_p^R$ to the partition functions of two R-replica statistical models with randomness.

Theorem 2. For the CSS code in the infinite system limit, $\text{Tr}\rho_p^R$ of the mixed state ρ_p corrupted by phase-flip errors is proportional to the partition function of the R-replica SM_2 with rRC. It is also proportional to the partition function of the R-replica SM_1 with iRC:

$$\text{Tr}\left(\rho_p^R\right) \propto \sum_{E \in \mathcal{V}} \left(Z_{\text{SM}_2}(\tilde{K}, E)\right)^R \propto \sum_{E \in \mathcal{V}} \left(W_{\text{SM}_1}(K, E)\right)^R, \quad (27)$$

where the coupling constants K and $\tilde{K}$ are given by the phase-flip error rate $p_z \in [0, 1/2]$ via $e^{-2\tilde{K}} = \tanh K = \frac{p_z}{1-p_z}$. The proportionality constants only depend on p_z smoothly and, hence, are unimportant for potential phases and DIPTs.

Here, the partition function of the R-replica SM_2 with rRC and that of the R-replica SM_1 with iRC are defined in parallel with Eqs. (23) and (25):

$$Z_{\text{SM}_2}(\tilde{K}, E) = \sum_{\{\bar{\tau}_j = \pm 1\}} \exp\left(\tilde{K} \sum_\mu (-1)^{E_\mu} \mathcal{O}_\mu^x[\tilde{\tau}]\right), \quad (28)$$

and

$$W_{\text{SM}_1}(K, E) = \sum_{\{\tau_i = \pm 1\}} \exp\left(-H_{\text{SM}_1}\right) \prod_\mu \left(\mathcal{O}_\mu^z[\tau]\right)^{E_\mu}$$

$$= \sum_{\{\tau_i = \pm 1\}} \exp\left(-H_{\text{SM}_1} - \frac{\mathbf{i}\pi}{2} \sum_\mu E_\mu \left(\mathcal{O}_\mu^z[\tau] - 1\right)\right). \quad (29)$$

Note that statistical models in Theorem 2 for the phase-flip errors are related to those in Theorem 1 for the bit-flip errors by interchanging the roles of rRC and iRC.

Similar to Eq. (24), we can show that

$$Z_{\text{SM}_2}(\tilde{K}, E) = [p_z(1 - p_z)]^{-N/2} \tilde{N}_s \sum_{C \in \mathcal{V}_z} P_z(E + C), \quad (30)$$

where $\tilde{N}_s$ is a factor associated with the internal symmetries of SM_2. This equation implies that $Z_{\text{SM}_2}(\tilde{K}, E)$ is the total probability of phase-flip error within the same equivalence class $[E]_z \in \mathcal{V}/\mathcal{V}_z$. All error chains within the same class $[E]_z$ only differ from each other by elements in $\mathcal{V}_z$ and, hence, lead to the same pattern of excitations on the $A_i[X]$ stabilizers.

Just like the case with bit-flip errors, the two R-replica random statistical models in Theorem 2 are dual to each other under an HLT duality, which we will discuss more in the next subsection.

With Eq. (30), we can take the $R \to 1$ limit of Theorem 2 and obtain the following corollary.

Corollary 2. The von Neumann entropy of the error-corrupted state ρ_p is given by the quenched-disorder-averaged free energy of SM_2 with rRC (up to unimportant additive constants):

$$S_1(\rho_p) = -\sum_{E \in \mathcal{V}} P_z(E) \log\left(Z_{\text{SM}_2}(\tilde{K}, E)\right), \quad (31)$$

where $\tilde{K}$ is given by the phase-flip error rate p_z via $e^{-2\tilde{K}} = \frac{p_z}{1-p_z}$.

Again, combining Ref. [8] and Eq. (31), we conclude that the phase-flip error decoding threshold of the code is the critical phase-flip error rate $p_z^\star(R)$ of the DIPT in the limit $R \to 1$.

C. Tapestry of dualities of a general CSS code

In this subsection, we discuss the tapestry (Fig. 1) woven by the dualities among the statistical models originating from a general CSS code $\mathcal{C}$.

First, there is a Kramers-Wannier-like high-low-temperature (HLT) duality between the nonrandom SM_1 and SM_2 obtained from the same CSS code $\mathcal{C}$ through ungauging. For bit-flip errors, Theorem 1 shows that two R-replica statistical models with random couplings describe the same quantity $\text{Tr}\rho_b^R$ associated with the Rth Rényi entropy $S_R(\rho_b)$ of the error-corrupted mixed state ρ_b. As shown in the bottom left corner of Fig. 1, these two statistical models are also dual to each other under an HLT duality, a generalization of the HLT duality that relates the nonrandom SM_1 and SM_2. Similarly, for phase-flip errors, the Rth Rényi entropy $S_R(\rho_p)$ of the mixed state ρ_p is described by another pair of R-replica statistical models with random couplings as stated in Theorem 2. These two R-replica statistical models are also dual under an HLT duality, as shown in the bottom right corner of Fig. 1. In addition to these HLT dualities, we find that there are extra dualities, dubbed the BPD (bit-phase-decoherence) dualities, that map the random R-replica statistical models associated with the bit-flip errors to those associated with phase-flip errors. In the following, we provide a general discussion and the physical intuition for these dualities. The full technical details of the proofs of these dualities are presented in the Appendices A and B.

Now, we discuss the HLT duality between the nonrandom SM_1 and SM_2. We provide below the physical intuition of this duality by comparing the high-temperature expansion of the partition of SM_1 and the low-temperature expansion of the partition function of SM_2. The partition function of the nonrandom SM_1 can be written as a high-temperature expansion:

$$
\begin{aligned}
Z_{SM_1}(K) &= \sum_{\{\tau_i = \pm 1\}} \exp\left(K \sum_\mu \mathcal{O}^z_\mu[\tau] \right) \\
&= \sum_{\{\tau_i = \pm 1\}} (\cosh K)^N \prod_\mu \left(1 + \mathcal{O}^z_\mu[\tau] \tanh K \right).
\end{aligned} \quad (32)
$$

Recall that N is the number of qubits in the original CSS code, which equals the number of $\mathcal{O}^z_\mu[\tau]$ terms in SM_1. After we expand the product over μ and sum over all the spin configurations $\{\tau_i = \pm 1\}$, the nonvanishing contribution must come from the products of $\mathcal{O}^z_\mu[\tau]$ terms that equal identity. Based on Eqs. (10) and (17), we know that $\prod_\mu (\mathcal{O}^z_\mu[\tau])^{(b_j)_\mu} = 1$ for any j. In fact, we can readily see that $\prod_\mu (\mathcal{O}^z_\mu[\tau])^{C_\mu} = 1$ for a local set $C \in \mathcal{V}$ implies $C \in \mathcal{V}_z$ (in the absence of any local logical operators in the code $\mathcal{C}$ which we have assumed). Hence, in the infinite system limit, we can write

$$
Z_{SM_1}(K) \propto (\cosh K)^N \sum_{C \in \mathcal{V}_z} (\tanh K)^{|C|}, \quad (33)
$$

with a proportionality constant only depending on the system size. Strictly speaking, the summation here should run over the space $\mathcal{V}_x^\perp$ instead. Note that the space $\mathcal{V}_x^\perp / \mathcal{V}_z$ is associated with the logical-Z operators of the code which are all nonlocal. Also, the number of logical-Z operators N_c obeys $N_c/N \to 0$ as $N \to \infty$. Hence, replacing $\sum_{C \in \mathcal{V}_x^\perp}$ by $\sum_{C \in \mathcal{V}_z}$ does not change the free energy density of SM_1 in the infinite system limit.

For SM_2, we perform a low-temperature expansion of the partition function. A classical ground state of the SM_2 Hamiltonian H_{SM_2} Eq. (20) is the state with $\tilde{\tau}_j = +1$ for all j. The ground-state energy is $-N\tilde{K}$. The energy cost of a single spin flip $\tilde{\tau}_j \to -1$ at the site j is $+2\tilde{K}|b_j|$. The low-temperature expansion is a summation over all possible spin flips on top of the classical ground state:

$$
Z_{SM_2}(\tilde{K}) \propto e^{N\tilde{K}} \sum_{C \in \mathcal{V}_z} e^{-2\tilde{K}|C|}, \quad (34)
$$

where the proportionality constant is given by the degeneracy of the classical Hamiltonian H_{SM_2}, which is independent of $\tilde{K}$ but related to the internal symmetry of H_{SM_2}. By comparing Eqs. (34) and (33), we establish the HLT duality between the nonrandom SM_1 and SM_2 under the condition

$$
\tanh K = e^{-2\tilde{K}}. \quad (35)
$$

This condition relates a large (small) positive value of K to a small (large) positive value of $\tilde{K}$, which is natural for an HLT duality. Another way to obtain this HLT duality is to apply Wegner's general approach for dualities in Ising-type models [25].

Now, we discuss the dualities among the random statistical models appearing in the decohered CSS code. In the case of bit-flip errors, as shown in Theorem 1, the R-replica SM_1 with

rRC is dual to the R-replica SM_2 with iRC under an HLT duality. The duality relation between the coupling constants K and $\tilde{K}$ is the same as Eq. (35) for the nonrandom case. The proof of the HLT duality between the two R-replica random statistical models generalizes the discussion above for the nonrandom case. The mathematical details of the proof are provided in Appendix A. For the case of phase-flip errors, a parallel analysis can be made for the two different R-replica random statistical models that describe $\mathrm{Tr}\rho_p^R$, as shown in Theorem 2. These two statistical models are also dual to each other under a similar HLT duality.

Next, we introduce the BPD dualities between the random statistical models associated with bit-flip errors and those associated with phase-flip errors. The statement of these BPD dualities is the following.

Theorem 3. For $R = 2, 3,$ and ∞, the R-replica random statistical models that describe the Rényi entropy $S_R(\rho_b)$ caused by the bit-flip errors are dual to the statistical models that describe the Rényi entropy $S_R(\rho_p)$ caused by the phase-flip errors. We can summarize these BPD dualities as

$$
\mathrm{Tr}(\rho_b^R) \propto \mathrm{Tr}(\rho_p^R), \quad \text{for } R = 2, 3, \text{ and } \infty \quad (36)
$$

(up to unimportant proportionality constants), when the error rates $0 < p_{x,z} < \frac{1}{2}$ satisfy the duality relations

$$
\left[(1 - p_x)^R + p_x^R\right]\left[(1 - p_z)^R + p_z^R\right] = \frac{1}{2^{R-1}}. \quad (37)
$$

In particular, in $R \to \infty$ limit, the duality condition reduces to

$$
(1 - p_x)(1 - p_z) = \tfrac{1}{2} \quad (38)
$$

for the range of error rates $p_{x,z} \in (0, 1/2)$ under consideration. Note that duality relation Eq. (37) maps the weak bit-flip decoherence (p_x close to 0) to the strong phase-flip decoherence (p_z close to $1/2$), and vice versa. Hence, the BPD duality is a "strong-weak" duality for the strength of decoherence. We provide the mathematical details of the proof of this duality in Appendix B. Here, we sketch the general idea that leads to these BPD dualities. Based on Theorems 1 and 2, we can use the R-replica SM_1 with rRC to describe $S_R(\rho_b)$ and the R-replica SM_1 with iRC to describe $S_R(\rho_p)$. Integrating out the randomness, rRC or iRC, results in the interactions between the R replicas of SM_1 (see Appendix C). For a general R, the interreplica interactions mediated by rRC differ from those mediated by iRC. However, the cases with $R = 2, 3$ are exceptions. The interreplica interactions mediated by rRC and iRC are identical when the error rates $p_{x,z}$ satisfy the relations Eq. (37). Therefore, there are BPD dualities between the R-replica random statistical models for $S_R(\rho_b)$ and $S_R(\rho_p)$ when $R = 2, 3$.

We can also understand the BPD duality with $R = 2$ using the HLT duality between the nonrandom SM_1 and SM_2. Combining Eqs. (13), (23), and (24), we can integrate out the randomness E (see Appendix C) and obtain

$$
\mathrm{Tr}\rho_b^2 \propto Z_{SM_1}(K') \quad \text{with } K' = \mathrm{arctanh}((1 - 2p_x)^2), \quad (39)
$$

where $Z_{SM_1}(K')$ is the partition function of the nonrandom SM_1. Essentially, integrating out the real random-coupling "renormalizes" the coupling constant from $K = -\frac{1}{2} \log \frac{p_x}{1-p_x}$

to K'. Similarly, we can show that

$$\text{Tr}\rho_{\text{p}}^2 \propto Z_{\text{SM}_2}(\tilde{K}') \quad \text{with } \tilde{K}' = \text{arctanh}((1-2p_z)^2). \quad (40)$$

The BPD duality with $R = 2$ shown in Theorem 3 is equivalent to the HLT duality between the nonrandom statistical models in Eqs. (39) and (40).

The BPD duality at $R \to \infty$ can also be justified using the HLT duality between the nonrandom SM_1 and SM_2. From Eq. (22), we notice that $\text{Tr}\rho_{\text{b}}^R|_{R\to\infty}$ is dominated by the randomness pattern $E \in \mathcal{V}$ that maximizes the partition function $Z_{\text{SM}_1}(K, E)$. A high-temperature expansion similar to Eq. (33) implies that the maximum of $Z_{\text{SM}_1}(K, E)$ can be reached by the trivial randomness pattern $E = 0$. Therefore, the behavior of $\text{Tr}\rho_{\text{b}}^R|_{R\to\infty}$ is dominated by (the Rth power of) the partition function $Z_{\text{SM}_1}(K)^R$ of the nonrandom SM_1. Similarly, $\text{Tr}\rho_{\text{p}}^R|_{R\to\infty}$ is dominated by the partition function $Z_{\text{SM}_2}(\tilde{K})^R$ without randomness. Therefore, the HLT duality between the nonrandom SM_1 and SM_2 leads to the BPD duality between the Rényi entropies $S_R(\rho_{\text{b}})$ and $S_R(\rho_{\text{p}})$ in the limit $R \to \infty$.

We comment that a technically similar approximate duality was discovered first in the spin glass context using the "replica trick" [26] and later applied to the error threshold of stabilizer codes [10,12] (at $R \to 1$ in our language). In contrast, the dualities at $R = 2, 3, \infty$ discussed in this paper are exact and are directly associated with the Renyi entropies S_R, which are physical quantities by themselves instead of models appearing in the intermediate steps of the replica trick.

D. Discussion on the interpretation of DIPTs and the monotonicity of critical error rates

In this subsection, we briefly discuss the physical interpretation of DIPTs with different values of R. Also, we present a conjecture on the monotonicity of the critical error rates $p_{x,z}^\star(R)$ of the DIPTs as functions of the Rényi index R.

As discussed in the previous subsections, the critical error rates $p_{x,z}^\star$ of the DIPT in the limit $R \to 1$ matches the error threshold for the decodability of the logical information in the error-corrupted CSS code. For the DIPT with $R = 2$, in addition to the singularity of the second Rényi entropy, we can also view it from the perspective of a quantum phase transition in the doubled Hilbert space, which we explain in the following.

The discussions of bit-flip errors and phase-flip errors are completely parallel. Let us take the former as an example. Using the Choi-Jamiołkowski isomorphism [27,28], we can map the mixed-state density matrix ρ_{b} to a pure state $|\rho_{\text{b}}\rangle\rangle$, called the Choi representation of ρ_{b}, in the doubled Hilbert space. $|\rho_{\text{b}}\rangle\rangle$ is related to its error-free counterpart $|\rho_0\rangle\rangle$ via

$$|\rho_{\text{b}}\rangle\rangle \propto \left(e^{\tilde{K}\sum_\mu X_\mu \otimes X_\mu}\right)|\rho_0\rangle\rangle, \quad (41)$$

with $\tilde{K} = \text{arctanh}(\frac{p_x}{1-p_x})$. One can show that $|\rho_{\text{b}}\rangle\rangle$ is a ground state of the following parent Hamiltonian in the doubled Hilbert space

$$H_{\text{b}}^D = H_{\text{css}} \otimes \mathbb{1} + \mathbb{1} \otimes H_{\text{css}} + 2\sum_j e^{-2\tilde{K}\sum_\mu (\text{b}_j)_\mu X_\mu \otimes X_\mu}, \quad (42)$$

where H_{css} is the CSS-code Hamiltonian given in Eq. (2). The construction of this parent Hamiltonian generalizes the

construction for the decohered cluster states studied in Ref. [20]. A similar construction for the decohered 2D toric code was given in Ref. [15]. Appendix F contains the details of the construction of Eq. (42) in the most general setting. An interesting property is that H_{b}^D is frustration-free up to additive constants. In other words, we can decompose this Hamiltonian into a sum of positive-semidefinite local terms, and each of such local terms annihilates the ground state $|\rho_{\text{b}}\rangle\rangle$.

Let us analyze the phase diagram of this model Eq. (42) in the doubled Hilbert space. When p_x is close to 0, we can treat the last term of H_{b}^D as a perturbation. Since H_{css} by itself is a gapped Hamiltonian, the ground state $|\rho_{\text{b}}\rangle\rangle$ for small p_x is smoothly connected to $|\rho_0\rangle\rangle$ which contains two decoupled copies of the same CSS code (in the doubled Hilbert space). Another regime is represented by the limit $p_x = 1/2$. The ground state at $p_x = 1/2$ is a stabilizer state with stabilizers,

$$A_i[X] \otimes \mathbb{1}, \ \mathbb{1} \otimes A_i[X], \ X_\mu \otimes X_\mu, \quad \text{and } B_j[Z] \otimes B_j[Z].$$

This stabilizer state at $p_x = 1/2$ is equivalent to a single copy of the original CSS code $\mathcal{C}$ (embedded inside the doubled Hilbert space). If two copies of the original CSS codes represent a quantum phase of matter different from a single copy, there must be a quantum phase transition between them. This quantum phase transition must occur at the critical error rate $p_x^\star(R = 2)$ of the DIPT with $R = 2$. Note that this quantum phase transition lives in the same spatial dimension as the statistical model but has nontrivial temporal dynamics in the doubled Hilbert space. The spatial correlation of this quantum phase transition can be equivalently captured by the random statistical models with $R = 2$ discussed in previous subsections. For example, the equal-time correlation function $\langle\langle\rho_{\text{b}}|O_1 \otimes O_2|\rho_{\text{b}}\rangle\rangle \propto \text{Tr}(\rho_{\text{b}}O_1\rho_{\text{b}}O_2^\mathsf{T})$ can be translated into a correlation function of the 2D random statistical models with $R = 2$. Here, O_2^T is the transpose of the operator O_2. When the statistical models for $R = 2$ exhibit critical behavior, the same critical behavior appears in the spatial correlation of the quantum system in the doubled Hilbert space, indicating a quantum phase transition.

The physical implications of the DIPTs with a general R (beyond the singularities of the Rényi entropies) and the relations among the DIPTs with different R's are both interesting future research directions. Pertaining to the latter, we present the following conjecture.

Conjecture: The critical error rates $p_{x,z}^\star(R)$ are both monotonically increasing functions of R.

Here, we have assumed that for every error type, bit-flip or phase-flip, and for every Rényi index R, there is a unique DIPT in the range of error rates $p_{x,z} \in (0, 1/2)$. The associated critical error rates are the subject of the conjecture above.

A piece of evidence for this conjecture is given by the relation

$$p_{x,z}^\star(R = 2) < p_{x,z}^\star(R \to \infty), \quad (43)$$

which we prove in the following. Take the case of bit-flip errors as an example. As discussed in Sec. II C, the behavior of S_R for both $R = 2$ and $R \to \infty$ are related to the nonrandom SM_1. The "renormalized" coupling constant in the $R = 2$ case is $K' = \text{arctanh}((1-2p_x)^2)$ [see Eq. (39)] and the coupling
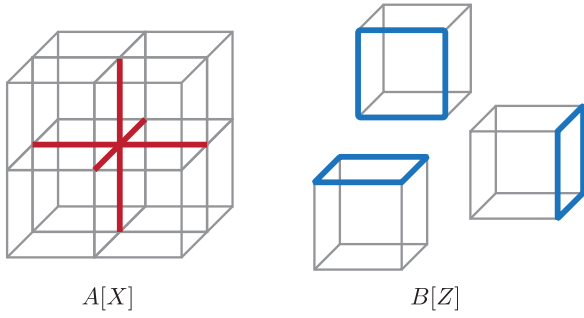
FIG. 3. Stabilizers of the 3D toric code. Qubits live on the edges of the lattice. An X-type stabilizer (red) is a product of 6 Pauli-X operators on the edges connected to the same site. A Z-type stabilizer (blue) is a product of 4 Pauli-Z operators on the edges of a plaquette.

constant in the $R \to \infty$ case is $K = \frac{1}{2} \log \frac{1-p_x}{p_x}$. The fact $K > K'$ for $p_x \in (0, \frac{1}{2})$ implies $p_x^\star(R = 2) < p_x^\star(R \to \infty)$.

As we will see in Sec. III, for the class of CSS codes with an *em* symmetry, the critical error rates $p_{x,z}^\star(R)$ for $R = 2, 3, \infty$ are exactly fixed [see Eq. (63)] by the dualities [assuming the uniqueness of DIPT for each R in the range $p_{x,z} \in (0, 1/2)$]. These exact critical error rates also obey our conjecture.

More evidence of our conjecture will be provided when we discuss specific examples, including the 3D toric code, the X-cube model, and the 2D toric code. In particular, we will see that our conjecture is consistent with the error thresholds, i.e., $p_{x,z}^\star(R \to 1)$, of these models obtained in previous literature.

If true, our conjecture establishes a general relation between DIPTs with different Rényi indices R. It would provide an interesting method to upper bound the error threshold $p_{x,z}^\star(R \to 1)$ using $p_{x,z}^\star(R = 2)$. Note that studying the limit $R \to 1$ requires the averaging over quenched disorders/randomness in the corresponding statistical models, while the $R = 2$ case is essentially captured by the same statistical model (with a renormalized coupling) but without randomness. We expect the latter to be generally simpler to analyze than the former.

Additionally, we remark that, since the statistical models for $R = 2, \infty$ can be mapped to the nonrandom SM$_1$ (see Sec. II C), the DIPTs at $R = 2$ and $R = \infty$ share the same universality class as the phase transition of the nonrandom SM$_1$ (when such phase transition exists).

E. Example: 3D toric code

As a pedagogical example, let us apply our formalism to the decohered toric code model. The 2D toric code belongs to a special class of CSS codes with an *em* symmetry, which we will focus on in Sec. III. Here, we consider the random statistical models that describe the 3D toric code with bit-flip and phase-flip errors.

The 3D toric code is a CSS code defined on a cubic lattice where the qubits, labeled by μ, are located on the edges of the lattice. As shown in Fig. 3, the 3D toric code has an X-type stabilizer $A_i[X]$ for every site i and a Z-type stabilizer $B_p[Z]$ for every plaquette p. Each $A_i[X]$ is a product of the Pauli-X operators on the six edges connected to the site i, while each
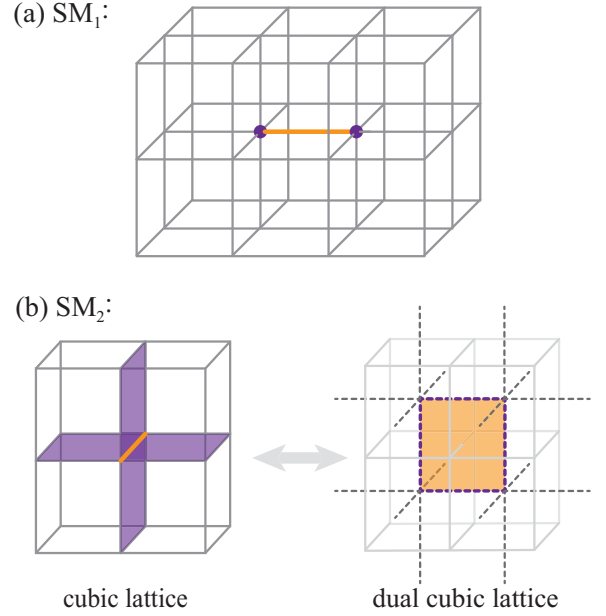

FIG. 4. Statistical models associated with the 3D toric code: (a) The 3D Ising model has a spin on each site of the cubic lattice and nearest-neighbor interactions (depicted as the two spins (purple dots) "coupled" by the orange edge). (b) Classical 3D $\mathbb{Z}_2$ gauge theory on the dual cubic lattice has a spin per dual edge (dashed lines). Each spin interaction involves the four classical spins on the boundary of a dual plaquette (orange).

$B_p[X]$ is a product of the Pauli-Z operators on the four edges that belong to the plaquette p:

$$A_i = \prod_{\mu \in \times_i} X_\mu, \qquad B_p = \prod_{\mu \in \Box_p} Z_\mu, \tag{44}$$

Now we construct SM$_1$ following the recipe in Sec. II B. We introduce a classical spin τ_i on each site i (where the $A_i[X]$ stabilizer is located). Equation (17) tells us that the spin interaction $\mathcal{O}_\mu^z$ is a product of τ_i's whose associated stabilizers A_i overlap with the edge μ. Hence, we have

$$\mathcal{O}_{\mu = \langle i, i' \rangle}^z = \tau_i \tau_{i'}, \tag{45}$$

where $\mu = \langle i, i' \rangle$ means μ is the edge between the nearest-neighbor pair of sites i and i' [see Fig. 4(a)]. The classical Hamiltonian of the nonrandom SM$_1$ is then given by

$$H_{\text{SM}_1} = -K \sum_{\langle i, i' \rangle} \tau_i \tau_{i'}, \tag{46}$$

which is exactly the Hamiltonian of the 3D classical Ising model (with the nearest-neighbor interaction).

For SM$_2$, we introduce a classical spin $\tilde{\tau}_p$ for each plaquette p. Each spin interaction $\mathcal{O}_\mu^x$ is the product of the four spins $\tilde{\tau}_p$ on the four plaquettes bordering the edge μ. Hence, the Hamiltonian of SM$_2$ is given by

$$H_{\text{SM}_2} = -\tilde{K} \sum_\mu \prod_{p \text{ s.t. } \mu \in \partial p} \tilde{\tau}_p, \tag{47}$$

where "$\prod_{p \text{ s.t. } \mu \in \partial p}$" represents the product over the plaquette p such that the edge μ belongs to the boundary ∂p of the

plaquette. As shown in Fig. 4(b), each spin interaction term in H_{SM_2}, when depicted on the dual cubic lattice, involves four classical spins on the four dual edges that form a dual plaquette. Hence, H_{SM_2} describes a 3D classical $\mathbb{Z}_2$ gauge theory on the dual lattice. It is well-known that the nonrandom 3D Ising model and the 3D classical $\mathbb{Z}_2$ gauge theory are dual to each other under an HLT duality (or Kramers-Wannier duality) [25].

In the following, we discuss the DIPTs in the decohered 3D toric code using the multireplica SM_1 and SM_2 with random couplings. We will focus on the cases with $R \to 1$ and $R = 2, 3, \infty$. Note that the rRC in the Ising model is also known as the random-bond disorder, while the rRC in the 3D $\mathbb{Z}_2$ gauge theory is often referred to as the random-plaquette disorder.

For $R \to 1$, it follows from Corollary 1 and 2 that, the von Neumann entropy $S_1(\rho_b)$ of the error-corrupted mixed state ρ_b is the quenched-disorder-averaged free energy of the 3D random-bond Ising model, while $S_1(\rho_p)$ is quenched-disorder-averaged free energy of the 3D random-plaquette $\mathbb{Z}_2$ gauge theory. The respective disorder distribution satisfies the Nishimori condition. As mentioned earlier, these quenched-disordered statistical models are exactly the ones previously used to study the bit-flip and phase-flip error thresholds $p_{x/z}^\star(R \to 1)$ for the 3D toric code [8,11,29]. The phase diagram of the 3D random-bond Ising model (in the $R \to 1$) was studied numerically in Refs. [30,31]. The 3D random-plaquette $\mathbb{Z}_2$ gauge theory was numerically investigated in Refs. [11,29]. These numerical studies obtained the following error thresholds for the 3D toric code:

$$p_x^\star(R \to 1) \approx 0.233, \quad p_z^\star(R \to 1) \approx 0.033. \quad (48)$$

For $R = 2$, as described around Eqs. (39) and (40), the two-replica random statistical models that describes the second Rényi entropy $S_2(\rho_{b/p})$ can be reduced to the nonrandom $SM_{1,2}$ with renormalized couplings $K' = \mathrm{arctanh}((1 - 2p_x)^2)$ and $\tilde{K}' = \mathrm{arctanh}((1 - 2p_z)^2)$. The critical point of SM_1, i.e., the 3D classical Ising model in this case, has been numerically studied (see Ref. [32], for example), which allows us to extract the critical bit-flip error rate $p_x^\star(2)$ for the DIPT with $R = 2$. The critical phase-flip error rate $p_z^\star(2)$ is related to the $p_x^\star(2)$ via the BPD duality Eq. (37). Therefore, we conclude

$$p_x^\star(2) \approx 0.266, \quad p_z^\star(2) \approx 0.099, \quad (49)$$

for the DIPTs with $R = 2$. These DIPTs belong to the 3D Ising universality class and, hence, share the same critical exponents. These transitions at $R = 2$ can also be understood from the perspective of the spontaneous breaking of $\mathbb{Z}_2$ higher-form symmetries. A counterpart with continuous higher-form symmetries was studied in Ref. [33].

For $R = 3$, the three-replica disordered statistical models that describe the behaviors of $S_3(\rho_{b/p})$ can be reduced to Askin-Teller-type statistical models, namely two coupled copies of $SM_{1,2}$, after the rRC or iRC are integrated out (see general discussions in Appendix C). For bit-flip errors, $S_3(\rho_b)$ is captured by the 3D Askin-Teller model on the cubic lattice:

$$H_{AT,b} = -K'' \sum_{\langle ii' \rangle} \left(\tau_i^{(1)} \tau_{i'}^{(1)} + \tau_i^{(2)} \tau_{i'}^{(2)} + \tau_i^{(1)} \tau_{i'}^{(1)} \tau_i^{(2)} \tau_{i'}^{(2)} \right). \quad (50)$$

TABLE I. Critical error rates of the DIPTs in the decohered 3D toric code.

R	$p_x^\star(R)$	$p_z^\star(R)$
1	0.233	0.033
2	0.266	0.099
3	0.288	0.135
∞	0.391	0.179

Here, the coupling constant K'' is given by the relation $\tanh K'' + (\tanh K'')^{-1} = (1 - 2p_x)^{-2} + 1$. $\tau_i^{(1)}$ and $\tau_i^{(2)}$ are the classical $\mathbb{Z}_2$ variables on the two copies of cubic lattices, respectively. Previous numerical studies showed that this model exhibits a first-order transition between the paramagnetic and the ferromagnetic phase at the critical coupling $K_c'' \approx 0.157$ [34]. This result implies that, with $R = 3$, the DIPT driven by bit-flip errors is a first-order transition occurring at the critical bit-flip error rate

$$p_x^\star(3) \approx 0.288. \quad (51)$$

For phase-flip errors, $S_3(\rho_p)$ is described by a 3D Askin-Teller-type $\mathbb{Z}_2$ gauge theory with the Hamiltonian

$$H_{AT,p} = -\tilde{K}'' \sum_\mu \left(\prod_{\mu \in \partial p} \tilde{\tau}_p^{(1)} + \prod_{\mu \in \partial p} \tilde{\tau}_p^{(2)} + \prod_{\mu \in \partial p} \tilde{\tau}_p^{(1)} \tilde{\tau}_p^{(2)} \right). \quad (52)$$

To the best of our knowledge, this model has not been studied before. The BPD duality implies that this model is dual to the model in Eq. (50). Therefore, the DIPT with $R = 3$ driven by phase-flip errors is also first order. It occurs at the critical phase-flip error rate

$$p_z^\star(3) \approx 0.135. \quad (53)$$

As shown in Sec. II C, in the limit $R \to \infty$, the behavior of the Rényi entropies $S_\infty(\rho_{b/p})$ are effectively described by the nonrandom $SM_{1,2}$ with the couplings $K = \frac{-1}{2} \log \frac{p_x}{1-p_x}$ and $\tilde{K} = \frac{-1}{2} \log \frac{p_z}{1-p_z}$. Hence, we can obtain the critical error rates:

$$p_x^\star(\infty) \approx 0.391, \quad p_z^\star(\infty) \approx 0.179. \quad (54)$$

The corresponding DIPTs belong to the 3D Ising universality class.

We have discussed the critical error rates of the DIPTs with several different R's in the decohered 3D toric code. The results are summarized in Table I. As we can see, these critical error rates agree with our conjecture that $p_{x,z}^\star(R)$ monotonically increase as functions of R.

F. Example: X-cube model

Our second example is the X-cube model, which is a CSS code that exhibits fracton topological order [18]. The X-cube model is defined on a 3D cubic lattice with a qubit located on each edge of the lattice. Each X-stabilizer involves four nearby edges forming a cross, while each Z-stabilizer involves
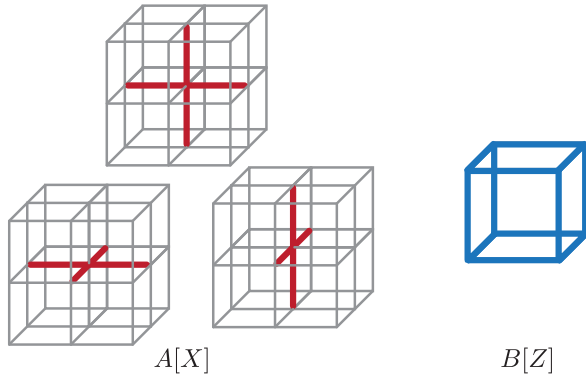
FIG. 5. X-type stabilizers (red) and Z-type stabilizers (blue) of the X-cube model.

12 edges of a unit cube (see Fig. 5):

$$A_{i,a} = \prod_{\mu \in +_{i,a}} X_\mu, \quad B_c = \prod_{\mu \in \boxdot_c} Z_\mu, \quad (55)$$

where $+_{i,a}$ denotes a cross formed by the four edges centered at site i, and $a = x, y, z$ labels the perpendicular direction to the cross. $\boxdot_c$ denotes the cube labeled by c.

In SM$_1$, there should be one classical spin per $A_{i,a}$ stabilizer. Hence, we introduce three colors of classical spins $\tau_{i,a=x,y,z}$ for each site i on the cubic lattice and represent them as the purple arrows in Fig. 6(a). The ungauging procedure in Sec. II B produces the following SM$_1$ Hamiltonian, which
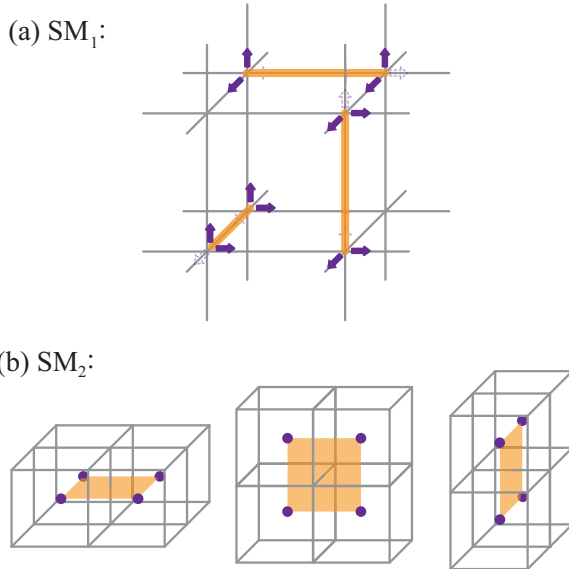

(a) SM$_1$:

(b) SM$_2$:

FIG. 6. Two statistical models originate from the X-cube model. (a) SM$_1$ is a model with three colors of classical spins (depicted by the purple arrows along the three directions) per site on a cubic lattice. Each nearest-neighbor interaction (orange) involves the four spins on the two neighboring sites whose colors (arrow directions) are perpendicular to the edge connecting the neighboring sites. (b) SM$_2$ is a plaquette Ising model on the dual cubic lattice. There is one classical spin per dual site. Each spin interaction term involves the four spins on the four corners of a dual plaquette.

describes a tricolor Ising model with the four-spin interactions

$$H_{\text{SM}_1} = -K \left(\sum_{\langle ii' \rangle \| \hat{z}} \tau_{i,x} \tau_{i,y} \tau_{i',x} \tau_{i',y} + \sum_{\langle ii' \rangle \| \hat{x}} \tau_{i,z} \tau_{i,y} \tau_{i',z} \tau_{i',y} \right.$$
$$\left. + \sum_{\langle ii' \rangle \| \hat{y}} \tau_{i,x} \tau_{i,z} \tau_{i',x} \tau_{i',z} \right). \quad (56)$$

For each nearest-neighbor pair of sites $\langle i, i' \rangle$, only the spins with their colors different from the direction of the edge $\langle i, i' \rangle$ are involved in the nearest-neighbor four-spin interaction [see Fig. 6(a)].

This tricolor Ising model has a classical gauge symmetry that changes $\tau_{i,a} \rightarrow -\tau_{i,a}$ independently for each site i. A change of variables effectively imposes a gauge fixing:

$$\eta_{i,x} \equiv \tau_{i,y} \tau_{i,z}, \quad \eta_{i,y} \equiv \tau_{i,x} \tau_{i,z}, \quad (57)$$

where $\eta_{i,x/y} = \pm 1$ are gauge invariant classical $\mathbb{Z}_2$ variables. Using these new variables, we can rewrite the Hamiltonian as an anisotropic 3D Ashkin-Teller model [10,35],

$$H_{\text{SM}_1} = -K \left(\sum_{\langle ii' \rangle \| \hat{z}} \eta_{i,x} \eta_{i,y} \eta_{i',x} \eta_{i',y} + \sum_{\langle ii' \rangle \| \hat{x}} \eta_{i,x} \eta_{i',x} \right.$$
$$\left. + \sum_{\langle ii' \rangle \| \hat{y}} \eta_{i,y} \eta_{i',y} \right). \quad (58)$$

Note we can also directly obtain this statistical model if we remove the $A_{i,z}$ stabilizers from the X-cube model. Removing the $A_{i,z}$ stabilizers does not change the nature of this CSS code because $A_{i,z}$ can be expressed using the remaining stabilizers, i.e., $A_{i,z} = A_{i,x} A_{i,y}$.

For SM$_2$, we introduce a classical spin $\tilde{\tau}_c$ to each cube c. The cube label c can also be viewed as the site index for the dual cubic lattice. The Hamiltonian of SM$_2$ is given by

$$H_{\text{SM}_2} = -\tilde{K} \sum_{\tilde{\Box}} \prod_{c \in \tilde{\Box}} \tilde{\tau}_c, \quad (59)$$

where $\tilde{\Box}$ sums over the dual plaquettes $\tilde{\Box}$ on the dual lattice and "$c \in \tilde{\Box}$" indicates that c is a dual site on one of the corners of $\tilde{\Box}$. H_{SM_2} represents a 3D plaquette Ising model with a four-spin interaction for each dual plaquette [see Fig. 6(b)].

We are now ready to examine the DIPTs of the X-cube model. The nonrandom anisotropic Ashkin-Teller model [Eq. (58)] and the 3D plaquette Ising model [Eq. (59)] were studied numerically in Ref. [36]. First-order phase transitions were found at the critical couplings $K_c \approx 0.657$ and $\tilde{K}_c \approx 0.276$. We can convert these results into the critical error rates of the DIPTs with $R = 2$ and $R = \infty$. The $R \rightarrow 1$ limit was numerically studied by Ref. [10] as the error thresholds of the X-cube model. We summarize these results in Table II. These numerically obtained critical error rates agree with our conjectured monotonicity of $p_{x,z}^\star(R)$.

TABLE II. Critical error rates for the X-cube model. The results for $R = 2, \infty$ are derived from the study of the (nonrandom) anisotropic Ashkin-Teller model and the plaquette Ising model in Ref. [36]. The error thresholds in the limit $R \to 1$ were obtained in Ref. [10].

R	$p_x^\star$	$p_z^\star$
1	0.075	0.152
2	0.120	0.241
∞	0.212	0.365

The $R = 3$ case requires the consideration of two coupled copies of plaquette Ising models on the dual cubic lattice:

$$H_{\mathrm{AT,p}} = -\tilde{K}'' \sum_{\tilde{\Box}} \left(\prod_{c \in \tilde{\Box}} \tilde{\tau}_c^{(1)} + \prod_{c \in \tilde{\Box}} \tilde{\tau}_c^{(2)} + \prod_{c \in \tilde{\Box}} \tilde{\tau}_c^{(1)} \tilde{\tau}_c^{(2)} \right) \quad (60)$$

and its BPD dual. To our knowledge, this model has not been studied before.

III. DECOHERED CSS CODE WITH ELECTRIC-MAGNETIC SYMMETRY

In this section, we study the consequences of the dualities for the CSS codes with an electric-magnetic (*em*) symmetry, a symmetry relating the X-type and Z-type stabilizers. We will see that the *em* symmetry effectively folds the tapestry of dualities in Fig. 1 in half, resulting in a new tapestry shown in Fig. 7. The new tapestry allows us to pin down the
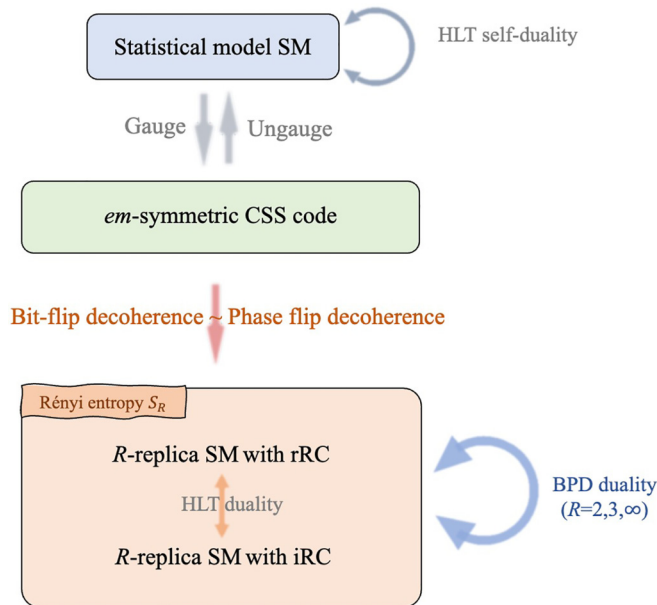

FIG. 7. The tapestry of dualities of an *em*-symmetric CSS code. The ungauging procedure leads to SM with a self-duality. Under bit-flip or phase-flip decoherence, R-replica SM with rRC and R-replica SM with iRC emerge as the description of the system's Rényi entropy. These random statistical models are related by HLT dualities for a general R. For $R = 2, 3, \infty$, the BPD dualities (combined with the *em* symmetry) relate different values of error rates.

super-universal self-dual error rates for the Rényi entropies $S_R(\rho_{\mathrm{b/p}})$ with $R = 2, 3, \infty$. These super-universal self-dual error rates must coincide with the critical error rates $p_{x/z}^\star$ of the DIPTs if there is a unique DIPT for each R. We will also discuss concrete examples of *em*-symmetric CSS codes, including the 2D toric code and Haah's code in 3D.

A. Tapestry of dualities and self-dual error rates in *em*-symmetric CSS codes

Let us first define the *em* symmetry. A CSS code is *em*-symmetric if there is a unitary transformation $\mathcal{U}_{em}$ such that

$$\mathcal{U}_{em} A_i[X] \mathcal{U}_{em}^\dagger = B_{j=f(i)}[Z],$$
$$\mathcal{U}_{em} B_j[X] \mathcal{U}_{em}^\dagger = A_{i=g(j)}[Z]. \quad (61)$$

Here, $\mathcal{U}_{em}$ is a product of the Hadamard gate $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ on each qubit, which exchanges the Pauli-X and the Pauli-Z operators, and a spatial action that permutes the locations of the qubits. The functions f and g are bijections between the locations of X- and Z-type stabilizers induced by the spatial action in $\mathcal{U}_{em}$.

Since the *em* symmetry maps between the X-type and Z-type stabilizers, the two statistical models $\mathrm{SM}_{1,2}$ produced by the ungauging procedure in Sec. II B become identical (up to some spatial rearrangement of the classical spins) in an *em*-symmetric CSS code. Hence, we will not distinguish $\mathrm{SM}_{1,2}$ in this section and call them both SM. The original HLT duality between the nonrandom SM_1 and SM_2 becomes an HLT self-duality of the nonrandom SM.

The *em* symmetry maps the bit-flip errors and the phase-flip errors into each other. Therefore, the decoherence effects caused by the two error types are effectively identical. We remark that the *em* symmetry is an intrinsic property of the CSS code under the current discussion. We do *not* require the error model to respect this symmetry. As followed from Theorems 1 and 2, the Rth Rényi entropies $S_R(\rho_{\mathrm{b/p}})$ are described by both the R-replica SM with rRC and R-replica SM with iRC. The two types of the R-replica random statistical models are dual to each other under the HLT dualities.

The BPD duality leads to profound consequences for *em*-symmetric CSS codes. Combining the *em* symmetry action and the BPD duality, both of which map between bit-flip errors and phase-flip errors, we conclude that the random statistical models for $S_R(\rho_{\mathrm{b}})$ with $R = 2, 3, \infty$ are self-dual, and so are those for $S_R(\rho_{\mathrm{p}})$.

Corollary 3. For an *em*-symmetric CSS code with decoherence, the Rényi entropies $S_R(\rho_{\mathrm{b/p}})$ with $R = 2, 3, \infty$ are described by R-replica random statistical models with self-dualities. The dualities map between the two error rates $0 < p, \tilde{p} < \frac{1}{2}$ related by

$$[(1-p)^R + p^R][(1-\tilde{p})^R + \tilde{p}^R] = \frac{1}{2^{R-1}}. \quad (62)$$

Here, both of the error rates p and $\tilde{p}$ can be either the bit-flip error rates or the phase-flip error rates, depending on the error type considered.

Similar to Theorem 3, the duality relations Eq. (62) are dualities between weak and strong decoherence strength. By solving the self-dual condition $p = \tilde{p}$ together with Eq. (62),
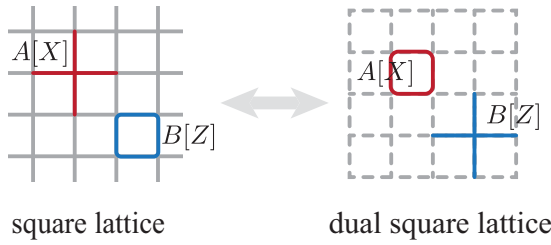
FIG. 8. The *em* symmetry of the 2D toric code is evident from the pictorial representations of the X- and Z-type stabilizers on the square lattice and its dual. The unitary operator $\mathcal{U}_{em}$ is a product of Hadamard gates on all sites and a translation action by half of a lattice spacing in both the horizontal and vertical directions.

we obtain the super-universal self-dual error rates (listed below) shared by *all em*-symmetric CSS codes. If we further assume that there is a unique DIPT for each R [within the range $p \in (0, 1/2)$], then the self-dual error rates must coincide with the critical error rates of the corresponding DIPTs:

$$
p^\star(R) = \begin{cases} \frac{1}{2}(1 - \sqrt{\sqrt{2} - 1}) \approx 0.178 & \text{for } R = 2, \\ \frac{1}{6}(3 - \sqrt{3}) \approx 0.211 & \text{for } R = 3, \\ \frac{1}{2}(2 - \sqrt{2}) \approx 0.293 & \text{for } R \to \infty. \end{cases} \tag{63}
$$

It is worth noting that these super-universal critical error rates are consistent with the conjectured monotonicity of $p^\star_{x/z}(R)$. Also, we called these values super-universal because they encompass DIPTs of different universality classes and in different dimensions, as demonstrated by the examples discussed below.

If there is more than one DIPT for $p \in [0, 1/2]$, the values given Eq. (63) should be viewed as the self-dual error rates. The critical error rates that differ from the self-dual error rates must form pairs. For every pair, there must be a critical error rate smaller than the self-dual value. Therefore, in the case of an *em*-symmetric CSS code with multiple DIPTs for $p \in [0, 1/2]$ and at a given R, the self-dual error rate in Eq. (63) upper bounds the critical error rate of the first DIPT encountered as the error rate increases from 0.

The tapestry summarizing all the dualities of an *em*-symmetric CSS code is shown in Fig. 7. Conceptually, this tapestry is Fig. 1 folded in half in the middle by the *em* symmetry.

In subsequent subsections, as demonstrations, we discuss two concrete models where our general results of *em*-symmetric CSS codes apply.

B. Example: 2D toric code

The 2D toric code is a familiar example of an *em*-symmetric CSS code. It is defined on a 2D square lattice with a qubit on each edge. As shown in Fig. 8, each X-type stabilizer is a product of four Pauli-X operators on the four edges forming a cross while each Z-type stabilizer is a product of four Pauli-Z operators on the four edges forming a plaquette:

$$
A_i = \prod_{\mu \in +_i} X_\mu \quad B_p = \prod_{\mu \in \square_p} Z_\mu, \tag{64}
$$

where $+_i$ labels the cross centered at the site i and $\square_p$ is the plaquette labeled by p.

It is straightforward to visualize the *em* symmetry by comparing the pictorial representations of the stabilizers on the square lattice and their representations on the dual square lattice (see Fig. 8). This *em* symmetry is generated by the unitary operator $\mathcal{U}_{em}$ that interchanges the X and Z operators on each qubit and translates the system both in the horizontal and vertical directions by half of the lattice spacing.

For the 2D toric code, SM is the 2D classical Ising model with the nearest-neighbor coupling:

$$
H_{\text{SM}} = -K \sum_{\langle i,i' \rangle} \tau_i \tau_{i'}, \tag{65}
$$

where the classical spins τ_i are located on the site of the square lattice. The HLT self-duality of this model (without randomness) is the well-known Kramers-Wannier duality.

Now, we discuss the Rényi entropies $S_R(\rho_{\text{b/p}})$ in the decohered 2D toric code. The $R \to 1$ limit is described by SM with quenched random-coupling disorder, which exactly recovers the random-bond Ising model introduced in Ref. [8] for the study of the error thresholds $p^\star(R \to 1)$. The cases with $R = 2$ and $R = \infty$ can be effectively captured by the nonrandom SM (see Appendix C). Hence, there is a unique DIPT for $R = 2, \infty$, whose universality class is equivalent to the Ising critical point in 2D. The corresponding critical error rates, fixed by the self-dualities, are given by Eq. (63).

For $R = 3$, integrating out the randomness in the statistical model yields the 2D square-lattice Ashkin-Teller model, whose Hamiltonian is given by

$$
H_{\text{AT}} = -K'' \sum_{\langle ii' \rangle} \left(\tau_i^{(1)} \tau_{i'}^{(1)} + \tau_i^{(2)} \tau_{i'}^{(2)} + \tau_i^{(1)} \tau_{i'}^{(1)} \tau_i^{(2)} \tau_{i'}^{(2)} \right). \tag{66}
$$

This Ashkin-Teller model tuned by the coupling constant K'' has a unique second-order transition [37], corresponding to $p^\star(3)$.

For $R = 4, 5, 6$, Ref. [14] numerically simulated the corresponding multireplica statistical models and found a unique phase transition in each case. The values of $p^\star(R)$ with $R \to 1$ and $R = 2, 3, 4, 5, 6, \infty$ are consistent with the conjectured monotonicity of $p^\star(R)$ as a function of R.

C. Example: Haah's code

Haah's code [19] provides an example of 3D *em*-symmetric CSS code. It is defined on a 3D cubic lattice with two qubits per site. There is one X-type and one Z-type stabilizer per cube. They are pictorially represented in Fig. 9(a). Each $A[X]$ stabilizer is a product of eight Pauli-X operators on the qubits in red, while each $B[Z]$ stabilizer is a product of eight Pauli-Z operators on the qubits in blue. The *em* symmetry action $\mathcal{U}_{em}$ of Haah's code is the product of (1) a Hadamard gate on each qubit, (2) a spatial inversion $\vec{r} \to -\vec{r}$, and (3) a swap of the two qubits on each site.

Apply the ungauging procedure in Sec. II B, we obtain a 3D SM with a classical $\mathbb{Z}_2$ spin per site on the dual cubic lattice. The Hamiltonian of this SM contains two types of four-spin interaction per unit cell, each associated with a tetrahedron inside a cube. They are pictorially represented in Fig. 9. This

(a) Haah's code:

$A[X]$ $B[Z]$
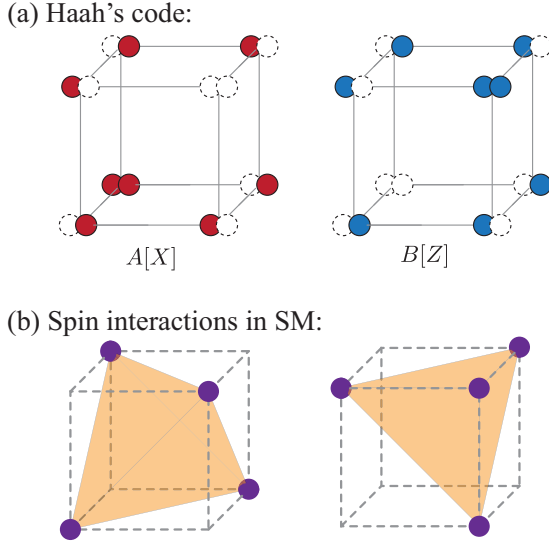
(b) Spin interactions in SM:

FIG. 9. (a) Haah's code is defined on a 3D cubic lattice with two qubits (depicted as the red, blue, and dashed circles) per site. For each cube, the X-type stabilizer is the product of the Pauli-X operators on the eight red qubits, and the Z-type stabilizer is the product of the Pauli-Z operators on the eight blue qubits. (b) In SM, which originates from Haah's code, there are two types of four-spin interactions in each unit cell. Each type is a product of the four classical spins (purple dots) on the corners of the corresponding tetrahedron (orange).

SM is also known as the "fractal Ising model." It was introduced in Ref. [18] as the classical model that produces Haah's code after gauging. The (nonrandom) fractal Ising model was recently shown through numerical studies to exhibit a unique first-order transition [38].

Following the recipe in Sec. II B, the Rényi entropies $S_R(\rho_{b/p})$ are mapped to the R-replica random fractal Ising models. As discussed earlier, the cases with $R = 2$ and $R \to \infty$ can be effectively reduced to the nonrandom fractal Ising model, which has a unique first-order transition. Therefore, the DIPTs with $R = 2$ and $R = \infty$ are first-order transitions occurring at the self-dual error rates listed in Eq. (63). For $R = 3$, the three-replica random fractal Ising model can be reduced to an Ashkin-Teller-type model after integrating out the randomness (see Appendix C). The phase diagram of this model and the nature of the self-dual error rate $p^\star(3)$ are both interesting problems that we leave for future investigation.

IV. GENERAL STABILIZER CODES UNDER DECOHERENCE

The mapping from the Rényi entropies to statistical models can be generalized to general stabilizer codes subject to a general Pauli noise channel (to be explained below). As we will see, we can associate a statistical model SM with a general stabilizer code without the CSS structure. The nonrandom version of SM has an HLT self-duality. The Rth Rényi entropy of the decohered stabilizer code is mapped to both the R-replica SM with rRC and the R-replica SM with iRC, which are related by an HLT duality. For $R = 2$ and $R \to \infty$, there is a general-Pauli-noise (GPN) duality that maps between two

different Pauli noise channels with different sets of error rates. We obtain the self-dual conditions of the Pauli noise channels and discuss their relations to DIPTs. We also include the Chamon model [39] as a concrete example to demonstrate our general construction.

A. Stabilizer codes and general Pauli noise

Let us start our discussion with the basics of stabilizer codes and the model for general Pauli noise. We still focus on the stabilizer code defined on a system of qubits (with each carrying a twofold local Hilbert space).

For a general stabilizer code, each stabilizer $A_J[X, Z]$ can be written as a product of the Pauli-X and Z operators. In this convention, any Pauli-Y operator in $A_J[X, Z]$ is decomposed into $\mathbf{i}XZ$. The subscript J of $A_J[X, Z]$ labels the stabilizer's center location (and the species of stabilizers if applicable). Each stabilizer $A_J[X, Z]$ can be represented by a pair of $\mathbb{Z}_2$ vectors $(\mathsf{a}_J, \mathsf{b}_J)$ with $\mathsf{a}_J, \mathsf{b}_J \in \mathcal{V}$:

$$A_J[X, Z] = (\mathbf{i})^{\mathsf{a}_J \cdot \mathsf{b}_J} \prod_\mu (X_\mu)^{(\mathsf{a}_J)_\mu} \prod_\mu (Z_\mu)^{(\mathsf{b}_J)_\mu}. \quad (67)$$

Recall $\mathcal{V} = \mathbb{Z}_2^N$, the $\mathbb{Z}_2$ vector space associated with a system of N qubits. The prefactor $(\mathbf{i})^{\mathsf{a}_J \cdot \mathsf{b}_J}$ ensures that $A_J[X, Z]^2 = \mathbb{1}$. The fact that all the stabilizers commute with each other is equivalent to the condition

$$\mathsf{a}_J \cdot \mathsf{b}_{J'} + \mathsf{a}_{J'} \cdot \mathsf{b}_J = 0, \quad \forall J, J'. \quad (68)$$

Note that the "·" above is the dot product in the $\mathbb{Z}_2$ vector space $\mathcal{V}$. This $\mathbb{Z}_2$-vector-based representation of the stabilizers is called "binary symplectic representation" in the literature [40]. Equation (68) defines the binary symplectic inner product between $(\mathsf{a}_J, \mathsf{b}_J)$ and $(\mathsf{a}_{J'}, \mathsf{b}_{J'})$. The binary symplectic representation is applicable to any Pauli strings. In other words, any product of Pauli operators can be represented by a pair of $\mathbb{Z}_2$ vectors $(\mathsf{e}, \mathsf{f}) \in \mathcal{V} \oplus \mathcal{V}$.

In this section, we consider the general Pauli noise that induces the Pauli noise channel $\mathcal{E} = \otimes_\mu \mathcal{E}_\mu$, a product of the local Pauli noise channel $\mathcal{E}_\mu$ on each qubit μ:

$$\begin{aligned} \mathcal{E}_\mu[\rho_0] &= (1 - p_x - p_y - p_z)\rho_0 + p_x X_\mu \rho_0 X_\mu \\ &\quad + p_y Y_\mu \rho_0 Y_\mu + p_z Z_\mu \rho_0 Z_\mu, \end{aligned} \quad (69)$$

where $p_{x,y,z} > 0$ are the probability/error rates of the X-, Y-, Z-type errors on each qubit. The general Pauli noise channel $\mathcal{E}$ includes the two quantum channels $\mathcal{N}_{x,z}$ induced by the bit-flip and the phase-flip errors and studied in Secs. II and III. $\mathcal{E}$ reduces to the depolarizing channel when we set $p_x = p_y = p_z$. For the following analysis, we make a technical assumption that $(1 - p_x - p_y - p_z) > p_{x,y,z}$ without the loss of generality. For a general Pauli noise channel, one can always combine it with an appropriate choice of unitary global action $\prod_\mu X_\mu$, $\prod_\mu Y_\mu$, or $\prod_\mu Z_\mu$ to construct a new quantum channel that satisfies the condition $(1 - p_x - p_y - p_z) > p_{x,y,z}$. Since this global action does not change the system's entropy, studying the decoherence effect of the new channel is equivalent to studying the original channel.

In what follows, we investigate the Rényi entropies $S_R(\rho_m)$ of the error-corrupted mixed state $\rho_m = \mathcal{E}(\rho_0)$ where ρ_0 is a pure logical state of the stabilizer code. Like the case of CSS

codes, we will construct the statistical models associated with $S_R(\rho_m)$ and study their dualities.

B. Statistical models for general stabilizer codes under general Pauli decoherence

To motivate the statistical model associated with a general decohered stabilizer code $\mathcal{C}$, it is useful first to consider the setting of CSS code with the general Pauli noise channel $\mathcal{E}$ defined above. In the case where the decoherence channel is generated by independent bit-flip and phase-flip errors, namely when $\mathcal{E} = \mathcal{N}_x \circ \mathcal{N}_z$, we have

$$p_x = (1 - q_z)q_x, \quad p_y = q_x q_z, \quad p_z = (1 - q_x)q_z, \quad (70)$$

where $q_{x/z}$ are the respective rates of the bit-flip and phase-flip errors in $\mathcal{N}_{x/z}$. The Y-type error, now occurring with probability $p_y = q_x q_z$, is the consequence of a bit-flip and a phase-flip error occurring on the same qubit. As discussed in Sec. II, the effects of independent bit-flip and phase-flip errors decouple on a CSS code. Therefore, the entropy of the entire system can be captured by two decoupled (random and multireplica) statistical models SM_1 and SM_2. The former lives on the lattice formed by the centers of the X-type stabilizers, while the latter lives on the lattice formed by the Z-type stabilizers.

It turns out that, when p_y deviates from $q_x q_z$, namely when the Y-type errors start to become independent from the bit-flip and phase-flip errors, SM_1 couples to SM_2 forming a single statistical model, which will be called SM in the following, living on the lattice formed by the centers of *all* stabilizers. The duality between SM_1 and SM_2 becomes the self-duality of SM. In fact, as we show below, even for a general stabilizer code $\mathcal{C}$ with Pauli decoherence, there is a corresponding self-dual SM whose R-replica random versions describe the Rényi entropies of the error-corrupted mixed state.

Like Sec. II, we first introduce the nonrandom version of the statistical model SM associated with the general stabilizer code $\mathcal{C}$. We assign one classical $\mathbb{Z}_2$ spin τ_J to each stabilizer $A_J[X, Z]$. The Hamiltonian of SM is given by

$$H_{SM} = -\sum_\mu K_x \mathcal{O}_\mu^x[\tau] + K_y \mathcal{O}_\mu^y[\tau] + K_z \mathcal{O}_\mu^z[\tau], \quad (71)$$

which contains three types of spin interactions

$$\mathcal{O}_\mu^z[\tau] = \prod_J (\tau_J)^{(\mathsf{a}_J)_\mu},$$

$$\mathcal{O}_\mu^x[\tau] = \prod_J (\tau_J)^{(\mathsf{b}_J)_\mu}, \quad (72)$$

$$\mathcal{O}_\mu^y[\tau] = \prod_J (\tau_J)^{(\mathsf{a}_J + \mathsf{b}_J)_\mu},$$

with coupling constants $K_{x,y,z}$. These spin interactions are the generalizations of Eqs. (17) and (21). They satisfy the relation

$$\mathcal{O}_\mu^x[\tau] \mathcal{O}_\mu^y[\tau] \mathcal{O}_\mu^z[\tau] = 1. \quad (73)$$

For the nonrandom SM, the comparison of the high-temperature and low-temperature expansion of the partition

function $Z_{SM} = \sum_{\{\tau_J = \pm 1\}} e^{-H_{SM}}$ yields a self-duality of SM that maps the coupling constants $\mathbf{K} = (K_x, K_y, K_z)$ to $\tilde{\mathbf{K}} = (\tilde{K}_x, \tilde{K}_y, \tilde{K}_z)$,

$$e^{-2(\tilde{K}_x + \tilde{K}_y)} = \frac{\tanh K_z + \tanh K_x \tanh K_y}{1 + \tanh K_x \tanh K_y \tanh K_z},$$

$$e^{-2(\tilde{K}_z + \tilde{K}_x)} = \frac{\tanh K_y + \tanh K_z \tanh K_x}{1 + \tanh K_x \tanh K_y \tanh K_z}, \quad (74)$$

$$e^{-2(\tilde{K}_y + \tilde{K}_z)} = \frac{\tanh K_x + \tanh K_y \tanh K_z}{1 + \tanh K_x \tanh K_y \tanh K_z}.$$

The detailed derivation of this duality in presented in Appendix D.

Under the Pauli noise channel, the error-corrupted mixed state of the stabilizer code is denoted as ρ_m. We prove the following theorem pertaining to the Rth Rényi entropy $S_R(\rho_m) = \frac{1}{1-R} \log \mathrm{Tr}(\rho_m^R)$ of the decohered stabilizer code.

Theorem 4. For the general stabilizer code in the infinite system limit, $\mathrm{Tr}\rho_m^R$ of the error-corrupted mixed state ρ_m generated by the Pauli noise channel is proportional to the partition function of the R-replica SM with the rRC. $\mathrm{Tr}\rho_m^R$ is also proportional to the partition function of the R-replica SM with iRC:

$$\mathrm{Tr}(\rho_m^R) \propto \sum_{E \in \mathcal{V} \oplus \mathcal{V}} (Z_{SM}(\mathbf{K}, E))^R \quad (75)$$

$$\propto \sum_{E \in \mathcal{V} \oplus \mathcal{V}} (W_{SM}(\tilde{\mathbf{K}}, E))^R, \quad (76)$$

with the coupling constants $\mathbf{K}$ and $\tilde{\mathbf{K}}$ given by the error rates $p_{x,y,z}$ via

$$e^{-2(K_x + K_y)} = \frac{\tanh \tilde{K}_z + \tanh \tilde{K}_x \tanh \tilde{K}_y}{1 + \tanh \tilde{K}_x \tanh \tilde{K}_y \tanh \tilde{K}_z}$$

$$= \frac{p_z}{1 - p_x - p_y - p_z},$$

$$e^{-2(K_y + K_z)} = \frac{\tanh \tilde{K}_x + \tanh \tilde{K}_y \tanh \tilde{K}_z}{1 + \tanh \tilde{K}_x \tanh \tilde{K}_y \tanh \tilde{K}_z}$$

$$= \frac{p_x}{1 - p_x - p_y - p_z},$$

$$e^{-2(K_x + K_z)} = \frac{\tanh \tilde{K}_y + \tanh \tilde{K}_x \tanh \tilde{K}_z}{1 + \tanh \tilde{K}_x \tanh \tilde{K}_y \tanh \tilde{K}_z}$$

$$= \frac{p_y}{1 - p_x - p_y - p_z}. \quad (77)$$

Here, $E \equiv (\mathsf{e}, \mathsf{f}) \in \mathcal{V} \oplus \mathcal{V}$ is the pair of $\mathbb{Z}_2$ vectors that represents the random couplings in the statistical models. We have defined the partition function of SM with the rRC pattern E (and the coupling constants $\mathbf{K}$) as

$$Z_{SM}(\mathbf{K}, E) = \sum_{\{\tau_J = \pm 1\}} \exp\left(\sum_\mu \left(K_x(-1)^{\mathsf{f}_\mu} \mathcal{O}_\mu^x[\tau] + K_y(-1)^{(\mathsf{e}+\mathsf{f})_\mu} \mathcal{O}_\mu^y[\tau] + K_z(-1)^{\mathsf{e}_\mu} \mathcal{O}_\mu^z[\tau]\right)\right), \quad (78)$$

and the SM partition function with the iRC pattern E (and the coupling constants $\tilde{\mathbf{K}}$) as

$$W_{\mathrm{SM}}(\tilde{\mathbf{K}}, E) = \sum_{\{\tau_I = \pm 1\}} \exp(-H_{\mathrm{SM}}(\tilde{\mathbf{K}})) \prod_\mu \left(\mathcal{O}_\mu^x[\tau]\right)^{\mathsf{e}_\mu} \left(\mathcal{O}_\mu^z[\tau]\right)^{\mathsf{f}_\mu}$$

$$= \sum_{\{\tau_I = \pm 1\}} \exp\left(-H_{\mathrm{SM}}(\tilde{\mathbf{K}}) - \frac{\mathbf{i}\pi}{2} \sum_\mu \mathsf{e}_\mu \left(\mathcal{O}_\mu^x[\tau] - 1\right) - \frac{\mathbf{i}\pi}{2} \sum_\mu \mathsf{f}_\mu \left(\mathcal{O}_\mu^z[\tau] - 1\right)\right). \tag{79}$$

The two types of R-replica random SM are dual to each other under the HLT duality. The proof of Theorem 4 (and the HLT duality) is provided in Appendix D.

The phase diagram of the decohered stabilizer code $\mathcal{C}$ has three tuning parameters $p_{x,y,z}$, which can potentially lead to even richer physics than the case of CSS codes with only bit-flip and phase-flip errors. The DIPTs of the decohered stabilizer code $\mathcal{C}$, indicated by the singularities in the Rényi entropy $S_R(\rho_{\mathrm{m}})$, corresponds to the phase transitions in these R-replica random statistical models.

Taking the limit of $R \to 1$ of Theorem 4, we obtain the relation between the von Neumann entropy $S_1(\rho_{\mathrm{m}})$ and the quenched-disorder-average partition function of SM with random couplings (derivation summarized in Appendix D):

Corollary 4. The von Neumann entropy of the error-corrupted state ρ_{m} is given by the quenched-disorder-averaged free energy of SM with rRC (up to an unimportant additive constant):

$$S_1(\rho_{\mathrm{m}}) = - \sum_{E \in \mathcal{V} \oplus \mathcal{V}} P(E) \log(Z_{\mathrm{SM}}(\mathbf{K}, E)), \tag{80}$$

with the generalized Nishimori condition Eq. (77) relating the probability distribution of the randomness and the coupling constant $\mathbf{K}$.

Here, the error chain $E \in \mathcal{V} \oplus \mathcal{V}$ represents a Pauli string in a similar fashion as Eq. (67). The probability $P(E)$ of the error chain is given by

$$P(E) = p_0^{N - N_x(E) - N_y(E) - N_z(E)} p_x^{N_x(E)} p_y^{N_y(E)} p_z^{N_z(E)}, \tag{81}$$

where $N_{x/y/z}(E)$ are the number of Pauli $X/Y/Z$ operators contained in the Pauli string E, $p_0 \equiv (1 - p_x - p_y - p_z)$, and N is the total number of qubits. Similar to the cases of CSS code, the $R \to 1$ limit recovers the statistical models introduced in the previous literature to specifically study the decodability and the error thresholds in stabilizer codes (see Ref. [12], for example).

At this point, we have finished the general construction of the random statistical models for $S_R(\rho_{\mathrm{m}})$. These statistical models offer tools to study emergent quantum matters in decohered stabilizer codes and the DIPTs between them. The physical implication of the DIPTs for a general R will be left for future studies.

For $R = 2$, based on the same reasoning as in Sec. II D, we can use the Choi-Jamiołkowski isomorphism to map the error-corrupted matrix state ρ_{m} to its Choi representation $|\rho_{\mathrm{m}}\rangle\rangle$ in the doubled Hilbert space. We can show that

$$|\rho_{\mathrm{m}}\rangle\rangle \propto \left(\prod_\mu e^{\tilde{K}_x X_\mu \otimes X_\mu - \tilde{K}_y Y_\mu \otimes Y_\mu + \tilde{K}_z Z_\mu \otimes Z_\mu}\right) |\rho_0\rangle\rangle, \tag{82}$$

where $\tilde{K}_{x,y,z}$ are given by $p_{x,y,z}$ through Eq. (77). The negative sign in front of the $Y_\mu \otimes Y_\mu$ terms is the consequence of $Y_\mu^{\mathsf{T}} = -Y_\mu$ under the Choi-Jamiołkowski isomorphism. The spatial correlation function on $|\rho_{\mathrm{m}}\rangle\rangle$ is identical to the correlation functions in the two-replica random SM. One can construct a frustration-free parent Hamiltonian similar to Eq. (42) for which $|\rho_{\mathrm{m}}\rangle\rangle$ is an exact ground state (see Appendix F for details). Therefore, the $R = 2$ DIPT for a decohered stabilizer code is related to a quantum phase transition in the doubled Hilbert space.

As a simple illustration of the general construction of the statistical models, we briefly discuss its application to the Chamon model [39] decohered by the Pauli noise channel. The Chamon model is defined on a cubic lattice with a single qubit per site. There is one stabilizer for each cube on the lattice. Each stabilizer is a product of six Pauli operators [see Fig. 10(a)]. The SM associated with the decohered Chamon model is defined on the dual cubic lattice with one classical $\mathbb{Z}_2$ spin per site. The interactions $\mathcal{O}^{x,y,z}$ in the classical Hamiltonian of SM are four-spin interactions illustrated in Fig. 10(b). To the best of our knowledge, the phase diagram of SM (with or without randomness) has not been studied before. It is interesting to study the possible phases of the decohered Chamon model using SM in the future.

(a) Stabilizers in Chamon model:
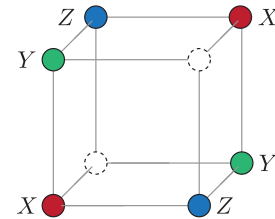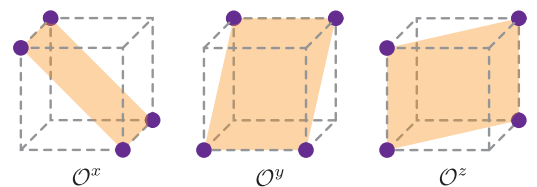

(b) Spin interactions in SM:

FIG. 10. (a) The stabilizer of Chamon model. (b) Three types of interaction $\mathcal{O}^{x,y,z}$ in SM: Each term (orange) is a product of four classical spins (purple dots). The dashed lines represents the edges of the dual cubic lattice.

C. Dualities for general decohered stabilizer codes

Now, we discuss the GPN (general Pauli noise) duality, a generalization of the BPD dualities (associated with the decohered CSS codes) to more general stabilizer codes. The statement of the GPN duality is the following.

Theorem 5. For $R = 2$, and ∞, the R-replica random statistical models that map to the Rényi entropy $S_R(\rho_m)$ at the error rates (p_x, p_y, p_z) are dual to those statistical models at the error rates $(\tilde{p}_x, \tilde{p}_y, \tilde{p}_z)$. The two sets of dual error rates obey

$$(1 - 2p_y - 2p_z)^2 = \frac{2\tilde{p}_0\tilde{p}_x + 2\tilde{p}_y\tilde{p}_z}{\tilde{p}_0^2 + \tilde{p}_x^2 + \tilde{p}_y^2 + \tilde{p}_z^2},$$

$$(1 - 2p_z - 2p_x)^2 = \frac{2\tilde{p}_0\tilde{p}_y + 2\tilde{p}_z\tilde{p}_x}{\tilde{p}_0^2 + \tilde{p}_x^2 + \tilde{p}_y^2 + \tilde{p}_z^2}, \quad \text{for } R = 2, \quad (83)$$

$$(1 - 2p_x - 2p_y)^2 = \frac{2\tilde{p}_0\tilde{p}_z + 2\tilde{p}_x\tilde{p}_y}{\tilde{p}_0^2 + \tilde{p}_x^2 + \tilde{p}_y^2 + \tilde{p}_z^2}.$$

and

$$\begin{aligned} 1 - 2p_y - 2p_z &= \tilde{p}_x/\tilde{p}_0, \\ 1 - 2p_z - 2p_x &= \tilde{p}_y/\tilde{p}_0, \quad \text{for } R = \infty, \quad (84) \\ 1 - 2p_x - 2p_y &= \tilde{p}_z/\tilde{p}_0. \end{aligned}$$

Here, $p_0 \equiv 1 - p_x - p_y - p_z$ and $\tilde{p}_0 \equiv 1 - \tilde{p}_x - \tilde{p}_y - \tilde{p}_z$.

A proof of this theorem is given in Appendix E. Conceptually, the GPN dualities with $R = 2$ and ∞ can be effectively viewed as the descendants of the self-duality of the nonrandom SM. For $R = 2$, after we integrate out the random couplings E, the two-replica random SM reduces to a single copy of nonrandom SM. For the limit $R = \infty$, the partition function of the R-replica random SM is dominated by the trivial random coupling configuration, i.e., $E = 0$. Therefore, the self-duality of the nonrandom SM implies the GPN dualities for $R = 2, \infty$. We caution that Theorem 4 is applicable only when the error rates $p_{x,y,z}$ and $\tilde{p}_{x,y,z}$ are both physical, namely they are between 0 and 1. In Appendix E, we also discuss how the GPN duality recovers the BPD duality for $R = 2, \infty$ for CSS codes.

For both $R = 2$ and $R = \infty$, there is a surface of self-dual error rates. The error rates (p_x, p_y, p_z) that satisfy the following equations map back to themselves under the GPN duality

$$\begin{aligned} (1 - p_x - p_y - p_z)^2 + p_x^2 + p_y^2 + p_z^2 &= \tfrac{1}{2}, \quad \text{for } R = 2, \\ p_x + p_y + p_z &= \tfrac{1}{2}, \quad \text{for } R = \infty. \end{aligned}$$
(85)

Note that these self-dual conditions are derived within the regime with $p_{x,y,z} < p_0 = 1 - p_x - p_y - p_z$. As commented earlier, the roles of p_0, p_x, p_y, and p_z can be permuted once the Pauli noise channel is followed by a global unitary action $\prod_\mu X_\mu$, $\prod_\mu Y_\mu$, or $\prod_\mu Z_\mu$. With these permutations taken into account, the surfaces of self-dual error rates are depicted in Fig. 11.

The surfaces of self-dual error rates are expected to constrain the phase diagram of the decohered stabilizer codes. One can argue that the renormalization group flow cannot cross this self-dual surface. A possible scenario is that this self-dual surface matches the critical surface of the DIPTs. The specific role of this self-dual surface in the phase diagram
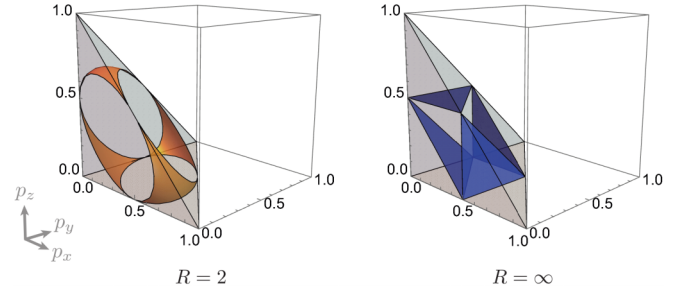

FIG. 11. The surfaces of self-dual error rates for $R = 2$ and $R = \infty$. The shaded area is the physical parameter regime with $p_{0,x,y,z} > 0$.

of the decohered stabilizer code should depend on the details of the model.

V. CONCLUSIONS AND OUTLOOK

In this paper, we systematically develop the formalism that studies the stabilizer codes decohered by Pauli noise/errors using statistical models. We focus on the Rényi entropies S_R of the decohered codes as a probe of the systems entanglement structure. We find a general mapping between the S_R's and classical statistical models that can be systematically constructed from the code's defining data. The phase transitions in these statistical models indicate nontrivial DIPTs in the decohered code. We discover intricate tapestries of dualities emerging among these statistical models. These dualities cast strong constraints on the phase diagram of the decohered quantum matter hosted in the stabilizer codes with noise. More specifically, this paper focuses on three general cases: (1) CSS codes decohered by bit-flip and phase-flip errors; (2) *em*-symmetric CSS codes under bit-flip and phase-flip decoherences; (3) general stabilizer codes under generic Pauli-noise decoherence.

First, for generic CSS codes, we construct a pair of statistical models SM_1 and SM_2, related to each other by an HLT duality [Eq. (35)], through an ungauging procedure. We show that the Rth Rényi entropy $S_R(\rho_b)$ for the error-corrupted mixed state ρ_b under the bit-flip decoherence is described by (1) R-replica SM_1 with rRC and (2) R-replica SM_2 with iRC. The two R-replica statistical models with different types of randomness are dual to each other through an HLT duality. Similarly, $S_R(\rho_p)$ for the error-corrupted mixed state ρ_p under phase-flip decoherence is described by (1) R-replica SM_2 with rRC and (2) R-replica SM_1 with iRC. These two models are also the HLT dual of each other. Moreover, for $R = 2, 3, \infty$, we find BPD dualities that relate $S_R(\rho_b)$ and $S_R(\rho_p)$, connecting the decoherence effects from the two error types. It is a "strong-weak" duality in that it maps strong bit-flip decoherence to weak phase-flip decoherence and vice versa. The tapestry of dualities associated with a general CSS code is illustrated in Fig. 1.

The classical statistical models that describe S_R provide us with powerful tools to investigate the phase diagrams of decohered CSS codes. In particular, the DIPTs in a decohered CSS code are identified as the phase transitions in the corresponding statistical models. The DIPTs with $R = 2, \infty$

are particularly simple as we have shown their equivalence to the phase transitions of $SM_{1,2}$ without randomness. In general, DIPTs for different R's happen at different critical error rates $p^\star_{x/z}(R)$. The universality classes with different R's are also different. Thus, there is a family of DIPTs indexed by R for a single CSS code. $p^\star_{x/z}(R \to 1)$ corresponds to the decoding error thresholds of the decohered CSS code, which has been extensively studied in many earlier works (for example, Refs. [8,11,29]). The DIPTs for $R \geqslant 2$ reveal other interesting singularities in the entanglement structure of the error-corrupted mixed states. For $R = 2$, the DIPT can also be interpreted as a quantum phase transition in a doubled Hilbert space. The nature and physical implication of the DIPTs at $p^\star(R)$ for $R \geqslant 3$ are interesting questions for the future. We also propose a conjecture on the monotonicity of the critical error rates $p^\star_{x/z}(R)$ as a function of R and present several pieces of evidence for this conjecture.

Second, for CSS codes with an *em* symmetry between the X-type and Z-type stabilizers, the corresponding SM_1 and SM_2 become the same model SM. The tapestry of dualities of *em*-symmetric CSS code becomes Fig. 7, which is effectively the tapestry of a general CSS code (Fig. 1) folded in half. The Rényi entropy S_R under either bit-flip or phase-flip decoherence is described by (1) R-replica SM with rRC and (2) R-replica SM with iRC. The two R-replica statistical models with iRC and rRC are related to each other through an HLT duality. The BPD duality for $R = 2, 3, \infty$ becomes a self-duality that relates strong bit-flip (phase-flip) decoherence to weak bit-flip (phase-flip) decoherence. Remarkably, these dualities yield super-universal self-dual error rates for $R = 2, 3, \infty$ [Eq. (63)]. If there is a unique DIPT (at a given R), the self-dual error rate must coincide with the critical error rate of the transition. In the two examples we examined, 2D toric code and Haah's code in 3D, the self-dual error rates indeed match the critical error rates of the DIPTs. The properties of the DIPTs in these two codes are very different, demonstrating that the self-dual error rates encompass different universality classes of phase transitions and are super-universal. For cases with multiple DIPTs for single R, the BPD self-duality implies that the DIPT away from the self-dual error rates must appear in pairs, and the self-dual error rates are upper bounds to the transition with the lowest error rate.

Finally, we extend our analysis to general stabilizer codes under decoherence of general Pauli noises, i.e., independent X-, Y-, Z-errors with error rates p_x, p_y, p_z. The description of the decohered code is centered around a single statistical model SM. For a CSS code, SM is the previously constructed $SM_{1,2}$ coupled together. The entropy $S_R(\rho_m)$ caused by the general Pauli noise is mapped to (1) R-replica SM with rRC and (2) R-replica SM with iRC. The two R-replica statistical models with randomness are related to each other by an HLT duality. In addition, we find a GPN self-duality that relates different error rates for $R = 2, \infty$, analogous to the BPD duality. Super-universal self-dual surfaces of error rates are identified for $R = 2, \infty$ (Fig. 11). These self-dual surfaces exist for a general stabilizer code and constrain the possible universal properties of the decohered code. One potential scenario is that the self-dual surfaces of error rates are exactly the phase boundary between different phases of the decohered code. A

possible next step of investigation is to understand the physical meaning of these self-dual surfaces in specific stabilizer codes.

Our systematic construction of the statistical models provides powerful tools to investigate the DIPTs of QEC codes. For future research, an important question to address concerns the quantum-information-theoretic interpretation of these DIPTs indexed by R. As explained, in the $R \to 1$ limit, the DIPT is associated with the decodability of the logical information from the decohered code. Does the DIPT with a given $R > 1$ mark the limit of decoherence for some concrete R-dependent quantum information processing protocol? Finding such protocols will deepen our understanding of the physical meaning of this family of DIPTs and will help refine the notion of topological order in error-corrupted mixed states. Moreover, such protocols will provide guidance on how to experimentally observe the DIPTs in a noisy intermediate-scale quantum platform. Additionally, it is an interesting future direction to establish more detailed relations between the critical behavior of statistical models at different R, including the critical correlation functions and scaling exponents, and the quantum-information-theoretic properties of the decohered code.

Another topic for future research pertains to our conjectured monotonic dependence of the critical error rates for DIPTs on the Rényi index R. The current conjecture concerns the bit-flip and phase-flip errors. If the conjecture is true, it will offer a tool to bound the critical error rates for a specific R, particularly the error threshold (in the limit of $R \to 1$), using results for larger R values. We have presented a few pieces of evidence of the conjecture. Developing a deeper understanding of the relationship between statistical models at different R will be important for resolving this conjecture. Moreover, it is also interesting to investigate if there is a similar relation between the critical error rates at different R for other types of errors, such as the general Pauli noise.

In this work, we have focused on the decoherence effects caused by Pauli noise/errors on the physical qubits of the QEC code. Investigating the effect of coherent errors, such as random-angle X or Z rotations, is a natural direction to generalize our framework. It was found that 2D surface code with coherent errors can be mapped to a 2D Ising model with complex couplings [41]. It would be interesting to generalize the systematic construction of the statistical models and the tapestry of dualities to general QEC codes with coherent errors. For a QEC code, another type of error that affects its performance is the measurement read-out error. For the decodability problem, taking the read-out error into account results in a statistical model living in one higher dimension than the spatial dimension of the QEC code [8]. An interesting open question is whether the measurement read-out errors can be investigated from the perspective of decohered quantum matter and if there is also a family of phase transitions similar to the family of DIPTs indexed by R.

Considering the decoherence-induced phases and transitions beyond stabilizer codes of qubit systems is another natural direction for future expeditions. For instance, it should be feasible to generalize our study to decohered stabilizer codes with qudits (each having a local Hilbert space of dimension $d > 2$). Exploring the decoherence effect in subsystems

codes [42,43] and the recently discovered Floquet codes [44] can potentially uncover new forms of decoherence-induced quantum matter.

Note added. While completing this manuscript, we noticed the independent related works [45,46].

APPENDIX A: PROOFS OF THEOREM 1 AND COROLLARY 1

Let ρ_b be the error-corrupted mixed state after bit-flip decoherence with error rate p. Combine Eqs. (8) and (14),

$$\text{Tr}\big(\rho_b^R\big) = \frac{(1-p)^{NR}}{2^{\dim \mathcal{V}_x}} \sum_{E \in \mathcal{V}} \sum_{C_{1,2,\dots R} \in \mathcal{V}_x} \lambda^{|E+C_1|} \cdots \lambda^{|E+C_R|}, \quad \text{(A1)}$$

where $\lambda = \frac{p}{1-p}$.

We will use two key observations. (1) the low-temperature expansion (LTE) of SM_1 generates $\mathcal{V}_x$; (2) the high-temperature expansion (HTE) of SM_2 generates $\mathcal{V}_z^\perp$ [see Eqs. (33) and (34)]. Take 3D toric code as an example (see Sec. II E). The corresponding SM_1 is the 3D Ising model, and SM_2 is the 3D lattice $\mathbb{Z}_2$ gauge theory. $\mathcal{V}_x$ consists of boundary (in the sense that it bounds a volume) surfaces on the dual lattice. LTE of Ising model gives domain walls which are precisely those surfaces, while HTE of lattice $\mathbb{Z}_2$ gauge theory gives closed surfaces (not necessarily a boundary) which corresponds to $\mathcal{V}_z^\perp$.

We will also substitute $\sum_{\mathcal{V}_z^\perp} \to \sum_{\mathcal{V}_x}$. This substitution essentially ignores the vectors in $\mathcal{V}_z^\perp / \mathcal{V}_x$, which corresponds to the logical operators. Under our assumption, there are no local logical operators. Hence, we argue that the contributions to the partition function from the vectors $\mathcal{V}_z^\perp / \mathcal{V}_x$ are exponentially suppressed as the system size goes to infinity. Additionally, we know that $N_c = \dim(\mathcal{V}_z^\perp / \mathcal{V}_x)$ is the number of logical qubits. And we have assumed the code rate $N_c/N \to 0$ in the large system limit. Hence, when we substitute $\sum_{\mathcal{V}_z^\perp}$ with $\sum_{\mathcal{V}_x}$, the change in the free energy density of the corresponding statistical model is negligible in the infinite-system limit.

Since free energy density is the standard diagnosis for phase transitions, we can do this substitution safely for the construction of the statistical models that target the DIPTs. In the following, we will use $\underline{N \to \infty}$ whenever this substitution happens.

We now have all the ingredients for the proof.

Proof of Theorem 1.

First consider LTE of SM_1,

$$Z_{\text{SM}_1}(K, E) = \sum_{\{\tau_i = \pm 1\}} \exp\left(K \sum_\mu (-1)^{E_\mu} \mathcal{O}_\mu^z[\tau]\right)$$

$$= N_s \sum_{C \in \mathcal{V}_x} \exp\left(K \sum_\mu (-1)^{E_\mu + C_\mu}\right)$$

$$= N_s \sum_{C \in \mathcal{V}_x} \exp(K(N - 2|E + C|))$$

$$= \lambda^{-N/2} N_s \sum_{C \in \mathcal{V}_x} \lambda^{|E+C|}. \quad \text{(A2)}$$

In the second line, the summation is over all possible "domain walls" C generated by spin flips. The interaction term flips the sign on the domain walls and takes the value $\mathcal{O}_\mu^z[\tau] = (-1)^{C_\mu}$. The symmetry factor N_s results from different spin flips giving the same domain walls when symmetry is present in the system. For instance, $N_s = 2$ for the Ising model because of the global $\mathbb{Z}_2$ symmetry. And $N_s = 3L$ for the plaquette Ising model, where L is the length of the lattice, due to a planar $\mathbb{Z}_2$ symmetry (flipping all the spins on any planes) [18]. In the third equality, we used the identity, $\sum_\mu (-1)^{E_\mu} = N - 2|E|$. Compared with Eq. (A1), we obtain

$$\text{Tr}\big(\rho_b^R\big) = \frac{[p(1-p)]^{RN/2}}{N_s^R 2^{\dim \mathcal{V}_x}} \sum_{E \in \mathcal{V}} \big(Z_{\text{SM}_1}(K, E)\big)^R. \quad \text{(A3)}$$

Second, consider HTE of SM_2,

$$W_{\text{SM}_2}(\tilde{K}, E) = \sum_{\{\tilde{\tau}_j = \pm 1\}} e^{\tilde{K} \sum_\nu \mathcal{O}_\nu^x[\tilde{\tau}]} \prod_\mu \big(\mathcal{O}_\mu^x[\tilde{\tau}]\big)^{E_\mu}$$

$$= (1 - \lambda^2)^{-N/2} \sum_{\{\tilde{\tau}_j = \pm 1\}} \prod_\nu \big(1 + \lambda \mathcal{O}_\nu^x[\tilde{\tau}]\big)$$

$$\times \prod_\mu \big(\mathcal{O}_\mu^x[\tilde{\tau}]\big)^{E_\mu}$$

$$= (1 - \lambda^2)^{-N/2} 2^{N_B} \sum_{C \in \mathcal{V}_z^\perp} \lambda^{|E+C|}$$

$$\overset{N \to \infty}{=\!=\!=} (1 - \lambda^2)^{-N/2} 2^{N_B} \sum_{C \in \mathcal{V}_x} \lambda^{|E+C|}. \quad \text{(A4)}$$

In the second equality, we used $\lambda = \tanh(\tilde{K})$. In the third equality, we kept the products of $\mathcal{O}_\nu^x$ which cancel all the spins in the random insertion $\prod_\mu (\mathcal{O}_\mu^x[\tilde{\tau}])^{E_\mu}$. Such products are necessarily of the form $\prod_\nu (\mathcal{O}_\nu^x[\tilde{\tau}])^{(E+C)_\nu}, C \in \mathcal{V}_z^\perp$. N_B is the number of $B_j[Z]$ stabilizers or number of spins $\tilde{\tau}_j$. Compared with Eq. (A1), we obtain

$$\text{Tr}\big(\rho_b^R\big) \overset{N \to \infty}{=\!=\!=} \frac{(1-2p)^{RN/2}}{2^{(\dim \mathcal{V}_x + N_B R)}} \sum_{E \in \mathcal{V}} \big(W_{\text{SM}_2}(\tilde{K}, E)\big)^R. \quad \text{(A5)}$$

As a side note, Eqs. (A2) and (A4) demonstrate the HLT duality between the random statistical models $Z_{\text{SM}_1}(K, E) \propto W_{\text{SM}_2}(\tilde{K}, E)$.

Finally, to obtain von Neumann entropy, combine Eqs. (15) and (A2):

$$S_1(\rho_b) = -\sum_{E \in \mathcal{V}} P_x(E) \log(Z_{\text{SM}_1}(K, E))$$

$$- \frac{N}{2} \log(p(1-p)) + \log(N_s). \quad \text{(A6)}$$

∎

APPENDIX B: PROOF OF THEOREM 3

Let us first consider bit-flip decoherence with error rate p and $\lambda = \frac{p}{1-p}$ (ignoring subscripts x for clarity). Start with Eq. (14),

$$\text{Tr}(\rho_b^R) = (1-p)^{RN} \sum_{E \in \mathcal{V}} \lambda^{|E|} \sum_{C_{2,\dots R} \in \mathcal{V}_x} \lambda^{|E+C_2|} \cdots \lambda^{|E+C_R|}. \tag{B1}$$

Now let us sum over E. Observe,

$$|E + C_\alpha| = \sum_{\mu=1}^{N} (E_\mu + C_{\alpha,\mu} - 2 E_\mu C_{\alpha,\mu}), \tag{B2}$$

where all arithmetic is done in $\mathbb{Z}$.

$$\text{Tr}(\rho_b^R) = (1-p)^{RN} \sum_{C_{2,\dots R} \in \mathcal{V}_x} \sum_{\{E_\mu=0,1\}} \lambda^{\sum_\mu [(R - 2\sum_{\alpha=2}^{R} C_{\alpha,\mu}) E_\mu + \sum_{\alpha=2}^{R} C_{\alpha,\mu}]}$$

$$= (1-p)^{RN} \sum_{C_{2,\dots R} \in \mathcal{V}_x} \prod_{\mu=1}^{N} \sum_{E_\mu=0,1} \lambda^{[(R-2n_\mu) E_\mu + n_\mu]}$$

$$= (1-p)^{RN} \sum_{C_{2,\dots R} \in \mathcal{V}_x} \prod_{\mu=1}^{N} (\lambda^{n_\mu} + \lambda^{R-n_\mu}) \tag{B3}$$

$$= (1-p)^{RN} (1 + \lambda^R)^N \sum_{C_{2,\dots R} \in \mathcal{V}_x} \prod_{\mu=1}^{N} \left(\frac{\lambda^{n_\mu} + \lambda^{R-n_\mu}}{1 + \lambda^R} \right), \tag{B4}$$

where we have defined the "occupation number" $n_\mu \equiv \sum_{\alpha=2}^{R} C_{\alpha,\mu}$. In the last step, we normalized the product so that unoccupied links contribute a weight of 1.

To prove the duality, we rearrange the sum in $\mathcal{V}_x$ into a sum in $\mathcal{V}_z$. Introduce an indicator function:
Lemma 1.

$$\delta(C \in \mathcal{V}_x) = \frac{1}{2^{\dim \mathcal{V}_x^\perp}} \sum_{\tilde{C} \in \mathcal{V}_x^\perp} (-1)^{C \cdot \tilde{C}}.$$

Proof. If $C \in \mathcal{V}_x$, then $C \cdot \tilde{C} = 0 \; \forall \tilde{C} \in \mathcal{V}_x^\perp$. Thus,

$$\text{RHS} = \frac{1}{2^{\dim \mathcal{V}_x^\perp}} \sum_{\tilde{C} \in \mathcal{V}_x^\perp} 1 = 1.$$

If $C \notin \mathcal{V}_x = (\mathcal{V}_x^\perp)^\perp$, then $\exists \tilde{C}_0 \in \mathcal{V}_x^\perp$, s.t. $C \cdot \tilde{C}_0 = 1$. We can rewrite

$$\text{RHS} = \frac{1}{2 \cdot 2^{\dim \mathcal{V}_x^\perp}} \sum_{\tilde{C} \in \mathcal{V}_x^\perp} \left((-1)^{C \cdot \tilde{C}} + (-1)^{C \cdot (\tilde{C} + \tilde{C}_0)} \right) = 0.$$

$\blacksquare$

Insert this indicator function into Eq. (B3),

$$\text{Tr}(\rho_b^R) = \frac{(1-p)^{RN}}{2^{(R-1)\dim \mathcal{V}_x^\perp}} \sum_{\tilde{C}_{2,\dots R} \in \mathcal{V}_x^\perp} \sum_{\{C_{\alpha,\mu}=0,1\}} \prod_{\mu=1}^{N} \left[(\lambda^{\sum_\alpha C_{\alpha,\mu}} + \lambda^{R - \sum_\alpha C_{\alpha,\mu}}) (-1)^{\sum_\alpha C_{\alpha,\mu} \tilde{C}_{\alpha,\mu}} \right]$$

$$= \frac{(1-p)^{RN}}{2^{(R-1)\dim \mathcal{V}_x^\perp}} \sum_{\tilde{C}_{2,\dots R} \in \mathcal{V}_x^\perp} \prod_{\mu=1}^{N} \left[\prod_{\alpha=2}^{R} \left(\sum_{C_{\alpha,\mu}=0,1} \lambda^{C_{\alpha,\mu}} (-1)^{C_{\alpha,\mu} \tilde{C}_{\alpha,\mu}} \right) + \lambda^R \prod_{\alpha=2}^{R} \left(\sum_{C_{\alpha,\mu}=0,1} \lambda^{-C_{\alpha,\mu}} (-1)^{C_{\alpha,\mu} \tilde{C}_{\alpha,\mu}} \right) \right]$$

$$= \frac{(1-p)^{RN}}{2^{(R-1)\dim \mathcal{V}_x^\perp}} \sum_{\tilde{C}_{2,\dots R} \in \mathcal{V}_x^\perp} \prod_{\mu=1}^{N} \left\{ (1+\lambda)^{R-1} [1 + (-1)^{\tilde{n}_\mu} \lambda] \left(\frac{1-\lambda}{1+\lambda} \right)^{\tilde{n}_\mu} \right\}$$

$$= \frac{1}{2^{(R-1)\dim \mathcal{V}_x^\perp}} \sum_{\tilde{C}_{2,\dots R} \in \mathcal{V}_x^\perp} \prod_{\mu=1}^{N} \left[\frac{1 + (-1)^{\tilde{n}_\mu} \lambda}{1 + \lambda} \left(\frac{1-\lambda}{1+\lambda} \right)^{\tilde{n}_\mu} \right]$$

$$\overset{N \to \infty}{=\!=\!=} \frac{1}{2^{(R-1)\dim \mathcal{V}_x^\perp}} \sum_{\tilde{C}_{2,\dots R} \in \mathcal{V}_z} \prod_{\mu=1}^{N} \left[\frac{1 + (-1)^{\tilde{n}_\mu} \lambda}{1 + \lambda} \left(\frac{1-\lambda}{1+\lambda} \right)^{\tilde{n}_\mu} \right], \tag{B5}$$

with $\tilde{n}_\mu \equiv \sum_{\alpha=2}^R \tilde{C}_{\alpha,\mu}$. In the last equality, we replaced $\mathcal{V}_x^\perp$ by $\mathcal{V}_z$. Now let us consider phase-flip error with error rate p_z, $\lambda_z \equiv \frac{p_z}{1-p_z}$. Applying Eq. (B4) but swapping X and Z,

$$\text{Tr}(\rho_p^R) = (1-p_z)^{RN}(1+\lambda_z^R)^N \sum_{\tilde{C}_{2,..R}\in\mathcal{V}_z} \prod_{\mu=1}^N \left(\frac{\lambda_z^{\tilde{n}_\mu} + \lambda_z^{R-\tilde{n}_\mu}}{1+\lambda_z^R} \right). \tag{B6}$$

Comparing Eqs. (B5) and (B6), the sum over $\mathcal{V}_z$ will have the same weight if we can match (restoring subscripts x),

$$\frac{1+(-1)^{\tilde{n}_\mu}\lambda_x}{1+\lambda_x}\left(\frac{1-\lambda_x}{1+\lambda_x}\right)^{\tilde{n}_\mu} = \frac{\lambda_z^{\tilde{n}_\mu} + \lambda_z^{R-\tilde{n}_\mu}}{1+\lambda_z^R}, \tag{B7}$$

for every $\tilde{n}_\mu = 1, \ldots, R-1$, with some function $\lambda_z(\lambda_x)$. Note λ_z represents the phase-flip error rate and shall not depend on $\tilde{n}$. We now show that this matching is possible for $R = 2, 3$.

For $R = 2$, Eq. (B7) gives only one equation for $\tilde{n} = 1$,

$$\left(\frac{1-\lambda_x}{1+\lambda_x}\right)^2 = \frac{2\lambda_z}{1+\lambda_z^2} \Leftrightarrow [(1-p_x)^2 + p_x^2][(1-p_z)^2 + p_z^2] = \frac{1}{2}. \tag{B8}$$

For $R = 3$, Eq. (B7) looks the same for $\tilde{n} = 1, 2$. There is still only one equation,

$$\left(\frac{1-\lambda_x}{1+\lambda_x}\right)^2 = \frac{\lambda_z + \lambda_z^2}{1+\lambda_z^3} \Leftrightarrow [(1-p_x)^3 + p_x^3][(1-p_z)^3 + p_z^3] = \frac{1}{4}. \tag{B9}$$

For $R \geqslant 4$, there are, in general, no solutions because the system of equations is overdetermined.

The $R \to \infty$ duality can be readily understood with the statistical models. We observe that the partition functions with randomness are always smaller than the nonrandom partition function:

Lemma 2.

$$Z_{\text{SM}_1}(K, E) \leqslant Z_{\text{SM}_1}(K), \quad W_{\text{SM}_1}(K, E) \leqslant Z_{\text{SM}_1}(K). \tag{B10}$$

Proof.
For the first part of the lemma, write $Z_{\text{SM}_1}(K, E)$ with HTE,

$$\begin{aligned}
Z_{\text{SM}_1}(K, E) &= \sum_{\{\tau_i=\pm 1\}} \exp\left(K \sum_\mu (-1)^{E_\mu} \mathcal{O}_\mu^z[\tau] \right) \\
&= \cosh(K)^N 2^{N_A} \sum_{\tilde{C}\in\mathcal{V}_x^\perp} \tanh(K)^{|\tilde{C}|}(-1)^{E\cdot\tilde{C}} \\
&\leqslant \cosh(K)^N 2^{N_A} \sum_{\tilde{C}\in\mathcal{V}_x^\perp} \tanh(K)^{|\tilde{C}|} = Z_{\text{SM}_1}(K),
\end{aligned} \tag{B11}$$

where N_A is the number of $A_i[X]$ stabilizers or number of τ_i spins.

For the second part of the lemma, write $W_{\text{SM}_1}(K, E)$ with LTE,

$$W_{\text{SM}_1}(K, E) = \sum_{\{\tau_i=\pm 1\}} e^{K\sum_\nu \mathcal{O}_\nu^z[\tau]} \prod_\mu (\mathcal{O}_\mu^z[\tau])^{E_\mu} = N_s \sum_{C\in\mathcal{V}_x} \exp(K(-1)^{|C|})(-1)^{C\cdot E} \leqslant N_s \sum_{C\in\mathcal{V}_x} \exp(K(-1)^{|C|}) = Z_{\text{SM}_1}(K). \tag{B12}$$

$\blacksquare$

Now consider a bit-flip error with $\frac{p_x}{1-p_x} = e^{-2K}$ and a phase-flip error with $\frac{p_z}{1-p_z} = \tanh(K)$. In both cases, the quantity $\text{Tr}(\rho^\infty) \approx (Z_{\text{SM}_1}(K))^R$ (up to proportionality constant), because the partition functions with $E \neq 0$ are exponentially suppressed by large R. Thus, the two decoherences are dual at $R \to \infty$ with the relation,

$$(1-p_x)(1-p_z) = \frac{1}{2}. \tag{B13}$$

Note we assumed $p_x, p_z \in (0, \frac{1}{2})$. We observe in all three cases $R = 2, 3, \infty$, the BPD duality relates strong bit-flip decoherence ($p_x \to 1/2$) to weak phase-flip decoherence ($p_z \to 0$) and vice versa.

APPENDIX C: INTEGRATE OUT THE RANDOMNESS OF REPLICA STATISTICAL MODELS

In this section, we integrate out randomness from the replica statistical models. We shall see that effective interreplica interactions emerge which are "mediated" by randomness. We also provide an understanding of the $R = 2, 3$ BPD dualities from

these statistical models. Consider bit-flip error-corrupted mixed state ρ_b and the quantity related to Rényi entropy $\text{Tr}(\rho_b^R)$. Let us first look at the corresponding statistical model with rRC. Combine Eqs. (14) and (A2) and the identity $\sum_\mu (-1)^{E_\mu} = N - 2|E|$,

$$\text{Tr}(\rho_b^R) \propto \sum_{E \in \mathcal{V}} \exp\left(K \sum_\mu (-1)^{E_\mu}\right) \left(Z_{\text{SM}_1}(K, E)\right)^{R-1}$$

$$= \sum_{\{E_\mu = 0,1\}} \sum_{\{\tau_j^\alpha = \pm 1\}} \exp\left(K \sum_\mu (-1)^{E_\mu}\left(1 + \sum_{\alpha=1}^{R-1} \mathcal{O}_\mu^z[\tau^\alpha]\right)\right)$$

$$= \sum_{\{\tau_j^\alpha = \pm 1\}} \prod_\mu \sum_{E_\mu = 0,1} \exp\left(K(-1)^{E_\mu}\left(1 + \sum_{\alpha=1}^{R-1} \mathcal{O}_\mu^z[\tau^\alpha]\right)\right)$$

$$= \sum_{\{\tau_j^\alpha = \pm 1\}} \prod_\mu \left(e^{K(1+\sum_{\alpha=1}^{R-1} \mathcal{O}_\mu^z[\tau^\alpha])} + e^{-K(1+\sum_{\alpha=1}^{R-1} \mathcal{O}_\mu^z[\tau^\alpha])}\right)$$

$$= \sum_{\{\tau_j^\alpha = \pm 1\}} \prod_\mu 2(\cosh(K))^R \sum_{r=0}^{R-1} \sum_{\alpha_1 < \alpha_2 < \ldots < \alpha_r} (\tanh(K))^r \left(\frac{1 + (-1)^r e^{-2K}}{1 + e^{-2K}}\right) \mathcal{O}^{\alpha_1} \ldots \mathcal{O}^{\alpha_r}, \quad \text{(C1)}$$

where in the last line we abbreviated $\mathcal{O}^\alpha \equiv \mathcal{O}_\mu^z[\tau^\alpha]$, $\alpha = 1, .., R-1$ is the replica index since we now have $R-1$ copies of spin models. The last line manifests arbitrary interreplica coupling $\mathcal{O}^{\alpha_1} \ldots \mathcal{O}^{\alpha_r}$ at the same site μ. We also define $\mathcal{O}^{\alpha_1} \ldots \mathcal{O}^{\alpha_r} = 1$ when $r = 0$.

Let us now consider the special cases $R = 2, 3$. First, when $R = 2$, the replica theory Eq. (C1) has only a single copy,

$$\text{Tr}(\rho_b^2) \propto \sum_{\{\tau_j = \pm 1\}} \prod_\mu \left(1 + (\tanh(K))^2 \mathcal{O}_\mu^z[\tau]\right) \propto \sum_{\{\tau_j = \pm 1\}} \prod_\mu e^{K' \mathcal{O}_\mu^z[\tau]} = Z_{\text{SM}_1}(K'), \quad \text{(C2)}$$

where $\tanh(K') = (\tanh(K))^2$. So upon integrating out errors E, the statistical model with randomness reduces to a clean model with renormalized coupling K'.

Second, when $R = 3$, there are now two replicas $\alpha = 1, 2$,

$$\text{Tr}(\rho_b^3) \propto \sum_{\{\tau_j^\alpha = \pm 1\}} \prod_\mu (1 + (\tanh(K))^2(\mathcal{O}^1 + \mathcal{O}^2 + \mathcal{O}^1\mathcal{O}^2)) \propto \sum_{\{\tau_j^\alpha = \pm 1\}} \prod_\mu e^{K''(\mathcal{O}^1 + \mathcal{O}^2 + \mathcal{O}^1\mathcal{O}^2)}, \quad \text{(C3)}$$

where $\frac{\tanh(K'') + (\tanh(K''))^2}{1 + (\tanh(K''))^3} = (\tanh(K))^2$. For Ising model, $\mathcal{O}_{\langle ij \rangle} = \tau_i \tau_j$, Eq. (C3) becomes the partition function of Ashkin-Teller model.

Let us now examine the case with iRC. Combine Eqs. (14) and (A4),

$$\text{Tr}(\rho_b^R) \propto \sum_{E \in \mathcal{V}} \exp\left(K \sum_\mu (-1)^{E_\mu}\right) \left(W_{\text{SM}_2}(\tilde{K}, E)\right)^{R-1}$$

$$= \sum_{\{E_\mu = 0,1\}} \sum_{\{\tilde{\tau}_j^\alpha = \pm 1\}} e^{K \sum_\mu (-1)^{E_\mu}} e^{\tilde{K} \sum_{v,\alpha} \mathcal{O}_v^x[\tilde{\tau}^\alpha]} \prod_{\sigma,\alpha} \left(\mathcal{O}_\sigma^x[\tilde{\tau}^\alpha]\right)^{E_\sigma}$$

$$= \sum_{\{\tilde{\tau}_j^\alpha = \pm 1\}} \prod_\mu e^{\tilde{K} \sum_\alpha \mathcal{O}_\mu^x[\tilde{\tau}^\alpha]} \sum_{E_\mu = 0,1} e^{K(-1)^{E_\mu}} \left(\prod_\alpha \mathcal{O}_\mu^x[\tilde{\tau}^\alpha]\right)^{E_\mu}$$

$$= \sum_{\{\tilde{\tau}_j^\alpha = \pm 1\}} \prod_\mu \frac{e^K}{\cosh(\tilde{K})} \exp\left(\tilde{K}\left(\sum_{\alpha=1}^{R-1} \tilde{\mathcal{O}}^\alpha + \prod_{\beta=1}^{R-1} \tilde{\mathcal{O}}^\beta\right)\right), \quad \text{(C4)}$$

where $\alpha, \beta = 1, .., R-1$ are the replica indices. In the last line, we again adopts the shorthand $\tilde{\mathcal{O}}^{\alpha,\beta} \equiv \mathcal{O}_\mu^x[\tilde{\tau}^{\alpha,\beta}]$. We see that the effective interaction in Eq. (C4) is more constrained than Eq. (C1). In the former, different replica copies only couple via the product $\prod_\alpha \tilde{\mathcal{O}}^\alpha$ of all copies, while in the latter, arbitrary products are present. However, this difference disappears when $R = 2, 3$, as we can observe by comparing Eqs. (C2)–(C4). This coincidence facilitates the BPD duality. We note that Eq. (C4) was derived for the special case of 2D toric code in Ref. [14] while our result applies for any CSS code.

APPENDIX D: PROOFS OF THEOREM 4 AND COROLLARY 4

Let us first introduce some notations. As mentioned in the main text, on an N-qubit system, a Pauli string E representing errors can be specified by a $\mathbb{Z}_2$ vector of length $2N$,

$$E = (\mathsf{e}, \mathsf{f}) \in \mathcal{V} \oplus \mathcal{V}. \tag{D1}$$

The Pauli operator acting on the μth qubit is $E_\mu = I, X, Y, Z$ if $(\mathsf{e}_\mu, \mathsf{f}_\mu) = (0, 0), (1, 0), (1, 1), (0, 1)$, respectively. In each error chain E, the number of X, Y, Z operators, $N_x(E), N_y(E), N_z(E)$ are given by

$$N_x(E) = \frac{|\mathsf{e}| - |\mathsf{f}| + |\mathsf{e} + \mathsf{f}|}{2},$$

$$N_y(E) = \frac{|\mathsf{e}| + |\mathsf{f}| - |\mathsf{e} + \mathsf{f}|}{2}, \tag{D2}$$

$$N_z(E) = \frac{-|\mathsf{e}| + |\mathsf{f}| + |\mathsf{e} + \mathsf{f}|}{2}.$$

Similarly, a stabilizer $A_J[X, Z]$ is represented by a vector

$$A_J = (\mathsf{a}_J, \mathsf{b}_J) \in \mathcal{V} \oplus \mathcal{V}. \tag{D3}$$

That all stabilizers commute amounts to $\mathsf{a}_J \cdot \mathsf{b}_{J'} + \mathsf{a}_{J'} \cdot \mathsf{b}_J = 0, \forall J, J'$. The set of stabilizers $\{A_J\}$ span a subspace $\mathcal{V}_s \subset \mathcal{V} \oplus \mathcal{V}$. We also define the dual subspace

$$\mathcal{V}_s^* \equiv \{\Omega W | W \in \mathcal{V}_s\}, \tag{D4}$$

where $\Omega = \left(\begin{smallmatrix} 0 & \mathbb{1}_{N \times N} \\ \mathbb{1}_{N \times N} & 0 \end{smallmatrix}\right)$. Note the matrix Ω effectively interchanges X and Z operators in a Pauli string. We can rewrite the commutation relation of stabilizers as

$$(A_J) \cdot (\Omega A_{J'}) = 0, \quad \forall J, J'. \tag{D5}$$

This implies $\mathcal{V}_s \subset (\mathcal{V}_s^*)^\perp$. Similar to the case of CSS code, the quotient space $(\mathcal{V}_s^*)^\perp / \mathcal{V}_s$ contains logical operators for the QEC code. We assume that the logical operators are nonlocal and the code rate $N_c/N \to 0$. As in the CSS code case, the contribution of such nonlocal operators to $\mathrm{Tr}(\rho^R)$ is exponentially small, and the overall correction to free energy density is negligible. We can, therefore, make the substitution $(\mathcal{V}_s^*)^\perp \to \mathcal{V}_s$ in what follows. As before, we will use $\stackrel{N \to \infty}{=\!=\!=}$ to signal this substitution.

Let us now see how the spaces $\mathcal{V}_s, \mathcal{V}_s^*$ emerge in the statistical model. Recall definitions of the interaction terms,

$$\mathcal{O}_\mu^z[\tau] = \prod_J (\tau_J)^{(\mathsf{a}_J)_\mu},$$

$$\mathcal{O}_\mu^x[\tau] = \prod_J (\tau_J)^{(\mathsf{b}_J)_\mu}, \tag{D6}$$

$$\mathcal{O}_\mu^y[\tau] = \mathcal{O}_\mu^x[\tau]\mathcal{O}_\mu^z[\tau].$$

LTE of this model gives the vectors in $\mathcal{V}_s$. To see this, consider a single flipped spin, $\tau_J = -1$ and $\tau_{J'} = 1, \forall J' \neq J$. The interaction terms become

$$\mathcal{O}_\mu^z[\tau] = (-1)^{(\mathsf{a}_J)_\mu}, \quad \mathcal{O}_\mu^x[\tau] = (-1)^{(\mathsf{b}_J)_\mu}. \tag{D7}$$

This domain wall pattern, therefore, encodes the vector $A_J = (\mathsf{a}_J, \mathsf{b}_J) \in \mathcal{V}_s$.

However, HTE of the statistical model yields the vectors in $(\mathcal{V}_s^*)^\perp$. The HTE sums over products of interaction terms that cancel all the spins. An arbitrary product of interaction terms has the form

$$\prod_\mu \left(\mathcal{O}_\mu^x[\tau]\right)^{\mathsf{a}_\mu} \left(\mathcal{O}_\mu^z[\tau]\right)^{\mathsf{b}_\mu}, \tag{D8}$$

where $C \equiv (\mathsf{a}, \mathsf{b}) \in \mathcal{V} \oplus \mathcal{V}$. To cancel all the spins, this combination of interactions must satisfy

$$1 = \prod_\mu \left(\mathcal{O}_\mu^x[\tau]\right)^{\mathsf{a}_\mu} \left(\mathcal{O}_\mu^z[\tau]\right)^{\mathsf{b}_\mu}$$

$$= \prod_\mu \prod_J (\tau_J)^{(\mathsf{b}_J)_\mu \mathsf{a}_\mu + (\mathsf{a}_J)_\mu \mathsf{b}_\mu}$$

$$= \prod_J (\tau_J)^{C \cdot (\Omega A_J)}. \tag{D9}$$

The exponent for each τ_J must be 0. This implies

$$C \cdot (\Omega A_J) = 0, \forall J \Leftrightarrow C \in (\mathcal{V}_s^*)^\perp. \tag{D10}$$

We observe here that, for SM, the space of LTE configurations $\mathcal{V}_s$ and the space of HTE configurations $(\mathcal{V}_s^*)^\perp$ are identical up to nonlocal terms. This hints at the HLT self-duality of SM, which will be evident in subsequent discussions.

Now consider the general decoherence $\rho_0 = |\Omega\rangle \langle\Omega| \to \rho_\mathrm{m}$,

$$\rho_\mathrm{m} = \sum_{E \in \mathcal{V} \oplus \mathcal{V}} P(E) O^\dagger(E) \rho_0 O(E),$$

$$P(E) = (1 - p_x - p_y - p_z)^{N - N_x - N_y - N_z} p_x^{N_x} p_y^{N_y} p_z^{N_z}, \tag{D11}$$

$$O(E) = (\mathbf{i})^{\mathsf{e} \cdot \mathsf{f}} \prod_\mu (X_\mu)^{\mathsf{e}_\mu} (Z_\mu)^{\mathsf{f}_\mu},$$

where $O(E)$ is the Pauli string specified by the error chain $E = (\mathsf{e}, \mathsf{f})$. For later convenience we rewrite the error rates with $\lambda_i \equiv \frac{p_i}{1 - p_x - p_y - p_z}$ for $i = x, y, z$,

$$P(E) = (1 - p_x - p_y - p_z)^N \lambda_x^{N_x} \lambda_y^{N_y} \lambda_z^{N_z} \tag{D12}$$

$$= (1 - p_x - p_y - p_z)^N \left(\sqrt{\frac{\lambda_y \lambda_z}{\lambda_x}}\right)^{|\mathsf{f}|} \left(\sqrt{\frac{\lambda_x \lambda_y}{\lambda_z}}\right)^{|\mathsf{e}|} \left(\sqrt{\frac{\lambda_x \lambda_z}{\lambda_y}}\right)^{|\mathsf{e} + \mathsf{f}|}. \tag{D13}$$

The information theoretical quantity that detects DIPTs is

$$\mathrm{Tr}\big(\rho_{\mathrm{m}}^{R}\big) = \sum_{E_{1,...R} \in \mathcal{V} \oplus \mathcal{V}} P(E_1)P(E_2)\dots P(E_R) \langle O(E_1)O^\dagger(E_2)\rangle_\Omega \dots \langle O(E_{R-1})O^\dagger(E_R)\rangle_\Omega \, \langle O(E_R)O^\dagger(E_1)\rangle_\Omega$$

$$= \sum_{E \in \mathcal{V} \oplus \mathcal{V}} \sum_{C_{2,...R} \in \mathcal{V}_s} P(E)P(E+C_2)\dots P(E+C_R) \langle O(E)O^\dagger(E+C_2)\rangle_\Omega \dots$$

$$\langle O(E+C_{R-1})O^\dagger(E+C_R)\rangle_\Omega \, \langle O(E+C_R)O^\dagger(E)\rangle_\Omega$$

$$= \sum_{E \in \mathcal{V} \oplus \mathcal{V}} \sum_{C_{2,...R} \in \mathcal{V}_s} P(E)P(E+C_2)\dots P(E+C_R)\langle O(E)O^\dagger(E+C_2)O(E+C_2)O^\dagger(E+C_3)\dots O(E+C_R)O^\dagger(E)\rangle_\Omega$$

$$= \sum_{E \in \mathcal{V} \oplus \mathcal{V}} \sum_{C_{2,...R} \in \mathcal{V}_s} P(E)P(E+C_2)\dots P(E+C_R) \tag{D14}$$

$$= \frac{1}{2^{\dim \mathcal{V}_s}} \sum_{E \in \mathcal{V} \oplus \mathcal{V}} \left[\sum_{C \in \mathcal{V}_s} P(E+C) \right]^R. \tag{D15}$$

In the second equality, we used the condition that $\langle OO^\dagger \rangle_\Omega \neq 0$ iff $OO^\dagger$ is generated by stabilizers. However, there is an ambiguity of powers of $\mathbf{i}$ in $\langle OO^\dagger \rangle_\Omega$ because the X, Y, Z's do not commute. To resolve this, in the third equality, we used the fact that $|\Omega\rangle$ is an eigenstate of all the operators $OO^\dagger$. In the fourth equality, we noted that the product of all $OO^\dagger$ is identity. In the last equality, we symmetrized the expression in a similar way as Eq. (14). It remains to relate $\sum_{C \in \mathcal{V}_s} P(E+C)$ to the partition functions of the statistical models with randomness.

First, consider LTE of SM with rRC $E = (\mathsf{e}, \mathsf{f}) \in \mathcal{V} \oplus \mathcal{V}$,

$$Z_{\mathrm{SM}}(\mathbf{K}, E) = \sum_{\{\tau_J = \pm 1\}} \exp\left(\sum_\mu \big(K_x(-1)^{\mathsf{f}_\mu}\mathcal{O}_\mu^x[\tau] + K_y(-1)^{(\mathsf{e}+\mathsf{f})_\mu}\mathcal{O}_\mu^y[\tau] + K_z(-1)^{\mathsf{e}_\mu}\mathcal{O}_\mu^z[\tau] \big) \right)$$

$$= N_s \sum_{C=(\mathsf{a},\mathsf{b}) \in \mathcal{V}_s} \exp\left(\sum_\mu \big(K_x(-1)^{(\mathsf{f}+\mathsf{b})_\mu} + K_y(-1)^{(\mathsf{e}+\mathsf{f}+\mathsf{a}+\mathsf{b})_\mu} + K_z(-1)^{(\mathsf{e}+\mathsf{a})_\mu} \big) \right)$$

$$= N_s e^{N(K_x+K_y+K_Z)} \sum_{C=(\mathsf{a},\mathsf{b}) \in \mathcal{V}_s} (e^{-2K_x})^{|\mathsf{f}+\mathsf{b}|}(e^{-2K_z})^{|\mathsf{e}+\mathsf{a}|}(e^{-2K_y})^{|(\mathsf{e}+\mathsf{a})+(\mathsf{f}+\mathsf{b})|}$$

$$= N_s e^{N(K_x+K_y+K_Z)} \sum_{C=(\mathsf{a},\mathsf{b}) \in \mathcal{V}_s} \left(\sqrt{\frac{\lambda_y\lambda_z}{\lambda_x}}\right)^{|\mathsf{f}+\mathsf{b}|} \left(\sqrt{\frac{\lambda_x\lambda_y}{\lambda_z}}\right)^{|\mathsf{e}+\mathsf{a}|} \left(\sqrt{\frac{\lambda_x\lambda_z}{\lambda_y}}\right)^{|(\mathsf{e}+\mathsf{a})+(\mathsf{f}+\mathsf{b})|}$$

$$= N_s [p_x p_y p_z (1 - p_x - p_y - p_z)]^{-N/4} \sum_{C \in \mathcal{V}_s} P(E+C), \tag{D16}$$

where N_s is a symmetry factor, in the third equality we used the identity $\sum_\mu (-1)^{\mathsf{w}_\mu} = N - 2|\mathsf{w}|$, $\forall \mathsf{w} \in \mathcal{V}$, in the fourth equality we applied generalized Nishimori condition Eq. (77), in the last equality we compared to Eq. (D13). Combine Eqs. (D15) and (D16),

$$\mathrm{Tr}\big(\rho_{\mathrm{m}}^{R}\big) = \frac{[p_x p_y p_z (1 - p_x - p_y - p_z)]^{NR/4}}{N_s^R 2^{\dim \mathcal{V}_s}} \sum_{E \in \mathcal{V} \oplus \mathcal{V}} (Z_{\mathrm{SM}}(\mathbf{K}, E))^R. \tag{D17}$$

This proves the first part of the theorem.

Now consider HTE of SM with iRC $E = (\mathsf{e}, \mathsf{f}) \in \mathcal{V} \oplus \mathcal{V}$,

$$W_{\mathrm{SM}}(\tilde{\mathbf{K}}, E) = \sum_{\{\tau_J = \pm 1\}} \exp\left(\sum_\mu \big(\tilde{K}_x\mathcal{O}_\mu^x[\tau] + \tilde{K}_y\mathcal{O}_\mu^y[\tau] + \tilde{K}_z\mathcal{O}_\mu^z[\tau] \big) \right) \prod_\nu \big(\mathcal{O}_\nu^x[\tau]\big)^{\mathsf{e}_\nu} \big(\mathcal{O}_\nu^z[\tau]\big)^{\mathsf{f}_\nu}$$

$$= f(\tilde{\mathbf{K}})^N \sum_{\{\tau_J\}} \prod_\mu \big(1 + \lambda_x\mathcal{O}_\mu^x[\tau] + \lambda_y\mathcal{O}_\mu^y[\tau] + \lambda_z\mathcal{O}_\mu^z[\tau]\big) \prod_\nu \big(\mathcal{O}_\nu^x[\tau]\big)^{\mathsf{e}_\nu} \big(\mathcal{O}_\nu^z[\tau]\big)^{\mathsf{f}_\nu}$$

$$= f(\tilde{\mathbf{K}})^N \sum_{\{\tau_J\}} \sum_{W=(\mathsf{a},\mathsf{b}) \in \mathcal{V} \oplus \mathcal{V}} \lambda_x^{N_x(W)}\lambda_y^{N_y(W)}\lambda_z^{N_z(W)} \prod_\mu \big(\mathcal{O}_\mu^x[\tau]\big)^{(\mathsf{a}+\mathsf{e})_\mu} \big(\mathcal{O}_\mu^z[\tau]\big)^{(\mathsf{b}+\mathsf{f})_\mu}$$

$$= f(\tilde{\mathbf{K}})^N 2^{N_A} \sum_{C \in (\mathcal{V}_s^*)^\perp} \lambda_x^{N_x(E+C)} \lambda_y^{N_y(E+C)} \lambda_z^{N_z(E+C)}$$

$$\stackrel{N \to \infty}{=\!=\!=} f(\tilde{\mathbf{K}})^N 2^{N_A} \sum_{C \in \mathcal{V}_s} \lambda_x^{N_x(E+C)} \lambda_y^{N_y(E+C)} \lambda_z^{N_z(E+C)}$$

$$= [(1 - 2p_x - 2p_y)(1 - 2p_x - 2p_z)(1 - 2p_y - 2p_z)]^{-N/4} 2^{N_A} \sum_{C \in \mathcal{V}_s} P(E+C), \tag{D18}$$

where $f(\tilde{\mathbf{K}}) = \cosh(\tilde{K}_x)\cosh(\tilde{K}_y)\cosh(\tilde{K}_z)(1 + \tanh(\tilde{K}_x)\tanh(\tilde{K}_y)\tanh(\tilde{K}_z))$ and N_A is the number of stabilizers $A_J[X, Z]$. In the second equality we expanded the exponential for each μ into a polynomial and used Eq. (77), in the third equality $W \in \mathcal{V} \oplus \mathcal{V}$ represents an arbitrary combination of interaction terms [cf. Eq. (D8)], in the fourth equality only $W = E + C, C \in (\mathcal{V}_s^*)^\perp$ survives the τ_J summation and the last equality applied Eq. (D12). Combine Eqs. (D15) and (D18),

$$\mathrm{Tr}(\rho_m^R) \stackrel{N \to \infty}{=\!=\!=} \frac{[(1 - 2p_x - 2p_y)(1 - 2p_x - 2p_z)(1 - 2p_y - 2p_z)]^{NR/4}}{2^{(\dim \mathcal{V}_s + N_A R)}} \sum_{E \in \mathcal{V} \oplus \mathcal{V}} (W_{\mathrm{SM}}(\tilde{\mathbf{K}}, E))^R. \tag{D19}$$

This completes the second part of the proof.

We now comment on the self-duality of the nonrandom SM [Eq. (74)]. Comparing Eqs. (D16) and (D18), we conclude that an HLT duality relates $Z_{\mathrm{SM}}(\mathbf{K}, E)$ and $W_{\mathrm{SM}}(\tilde{\mathbf{K}}, E)$ for arbitrary E. Setting $E = 0$ gives the HLT self-duality of nonrandom SM, $Z_{\mathrm{SM}}(\mathbf{K}) \propto Z_{\mathrm{SM}}(\tilde{\mathbf{K}})$.

Finally, to obtain the von Neumann entropy $S_1(\rho_m)$, combine Eqs. (D14) and (D16) and take the $R \to 1$ limit for S_R,

$$S_1(\rho_m) = - \sum_{E \in \mathcal{V} \oplus \mathcal{V}} P(E) \log(Z_{\mathrm{SM}}(\mathbf{K}, E)) - \frac{N}{4} \log(p_x p_y p_z (1 - p_x - p_y - p_z)) + \log(N_s). \tag{D20}$$

APPENDIX E: PROOF OF THE GPN DUALITY (THEOREM 5)

Throughout this Appendix, we use the shorthands,

$$p_0 \equiv 1 - p_x - p_y - p_z, \quad \lambda_i \equiv \frac{p_i}{p_0} \quad \text{for } i = x, y, z. \tag{E1}$$

1. Derivation for $R = 2$

Purity $\mathrm{Tr}(\rho_m^2)$ can be written as

$$\sum_{E_{1,2} \in \mathcal{V} \oplus \mathcal{V}} P(E_1) P(E_2) \delta(E_1 + E_2 \in \mathcal{V}_s). \tag{E2}$$

There are two ways to manipulate the $\delta(E_1 + E_2 \in \mathcal{V}_s)$. First, one can write

$$\sum_{E_{1,2} \in \mathcal{V} \oplus \mathcal{V}} P(E_1) P(E_2) \delta(E_1 + E_2 \in \mathcal{V}_s) = \sum_{E \in \mathcal{V} \oplus \mathcal{V}} \sum_{C \in \mathcal{V}_s} P(E) P(E + C)$$

$$= p_0^{2N} \sum_{C \in \mathcal{V}_s} \left(\sum_{E \in \mathcal{V} \oplus \mathcal{V}} \lambda_x^{N_x(E)+N_x(E+C)} \lambda_y^{N_y(E)+N_y(E+C)} \lambda_z^{N_z(E)+N_z(E+C)} \right)$$

$$\equiv p_0^{2N} \sum_{C \in \mathcal{V}_s} \prod_\mu f(C_\mu, \boldsymbol{p}). \tag{E3}$$

The sum on E is unconstrained and can be performed independently on each qubit with $E_\mu = I, X, Y, Z$, giving a Boltzmann weight,

$$f(C_\mu, \boldsymbol{p}) = \begin{pmatrix} 1 + \lambda_x^2 + \lambda_y^2 + \lambda_z^2 \\ 2\lambda_x + 2\lambda_y \lambda_z \\ 2\lambda_y + 2\lambda_x \lambda_z \\ 2\lambda_z + 2\lambda_x \lambda_y \end{pmatrix} = \frac{1}{p_0^2} \begin{pmatrix} p_0^2 + p_x^2 + p_y^2 + p_z^2 \\ 2p_x p_0 + 2p_y p_z \\ 2p_y p_0 + 2p_x p_z \\ 2p_z p_0 + 2p_x p_y \end{pmatrix}, \tag{E4}$$

where the four rows corresponds to $C_\mu = I, X, Y, Z$.

The second way to write $\delta(E_1 + E_2 \in V)$ is through the following resolution:

$$\delta(E_1 + E_2 \in \mathcal{V}_s) = \frac{1}{2^{\dim(\mathcal{V}_s^*)^\perp}} \sum_{C \in (\mathcal{V}_s^*)^\perp} (-1)^{C \cdot \Omega(E_1 + E_2)}. \tag{E5}$$

Purity now becomes

$$\mathrm{Tr}\big(\rho_{\mathrm{m}}^2\big) = \frac{1}{2^{\dim(\mathcal{V}_s^*)^\perp}} \sum_{C \in (\mathcal{V}_s^*)^\perp} \sum_{E_{1,2} \in \mathcal{V} \oplus \mathcal{V}} P(E_1) P(E_2)(-1)^{C \cdot \Omega(E_1 + E_2)}$$

$$= \frac{1}{2^{\dim(\mathcal{V}_s^*)^\perp}} \sum_{C \in (\mathcal{V}_s^*)^\perp} \left(\sum_{E \in \mathcal{V} \oplus \mathcal{V}} P(E)(-1)^{C \cdot (\Omega E)} \right)^2$$

$$\overset{N \to \infty}{=\!=\!=} \frac{(1 - \sum_i p_i)^{2N}}{2^{\dim(\mathcal{V}_s^*)^\perp}} \sum_{C \in \mathcal{V}_s} \left(\sum_{E \in \mathcal{V} \oplus \mathcal{V}} \lambda_x^{N_x(E)} \lambda_y^{N_y(E)} \lambda_z^{N_z(E)} (-1)^{C \cdot (\Omega E)} \right)^2$$

$$\equiv \frac{(1 - \sum_i p_i)^{2N}}{2^{\dim(\mathcal{V}_s^*)^\perp}} \sum_{C \in \mathcal{V}_s} \prod_\mu \tilde{f}(C_\mu, \boldsymbol{p}). \tag{E6}$$

Again E can be summed for each link individually $E_\mu = I, X, Y, Z$. Note $(-1)^{C \cdot (\Omega E)}$ tells us to insert a (-1) whenever $[C_\mu, E_\mu] \neq 0$. This gives Boltzmann weight,

$$\tilde{f}(C_\mu, \boldsymbol{p}) = \begin{pmatrix} (1 + \lambda_x + \lambda_y + \lambda_z)^2 \\ (1 + \lambda_x - \lambda_y - \lambda_z)^2 \\ (1 + \lambda_y - \lambda_x - \lambda_z)^2 \\ (1 + \lambda_z - \lambda_x - \lambda_y)^2 \end{pmatrix}$$

$$= \frac{1}{p_0^2} \begin{pmatrix} 1 \\ (1 - 2p_y - 2p_z)^2 \\ (1 - 2p_x - 2p_z)^2 \\ (1 - 2p_x - 2p_y)^2 \end{pmatrix}. \tag{E7}$$

Equating the normalized Boltzmann weights,

$$\frac{f(C_\mu, \boldsymbol{p})}{f(I, \boldsymbol{p})} = \frac{\tilde{f}(C_\mu, \tilde{\boldsymbol{p}})}{\tilde{f}(I, \tilde{\boldsymbol{p}})}, \tag{E8}$$

for $C_\mu = X, Y, Z$ gives the GPN duality at $R = 2$.

2. Derivation for $R = \infty$

At $R \to \infty$ the summation over randomness is dominated by the clean limit $E = 0$,

$$0 \leqslant Z(\mathbf{K}, E) \leqslant Z(\mathbf{K}), \quad 0 \leqslant W(\mathbf{K}, E) \leqslant Z(\mathbf{K}). \tag{E9}$$

But $Z(\mathbf{K}, E)$ and $W(\mathbf{K}, E)$ describe decoherence channels with error rates $\boldsymbol{p}$ and $\tilde{\boldsymbol{p}}$, respectively,

$$\lambda_x = e^{-2(K_y + K_z)}, \quad \lambda_y = e^{-2(K_x + K_z)}, \quad \lambda_z = e^{-2(K_x + K_y)},$$

$$\tilde{\lambda}_x = e^{-2(\tilde{K}_y + \tilde{K}_z)}, \quad \tilde{\lambda}_y = e^{-2(\tilde{K}_x + \tilde{K}_z)}, \quad \tilde{\lambda}_z = e^{-2(\tilde{K}_x + \tilde{K}_y)}, \tag{E10}$$

where $\mathbf{K}, \tilde{\mathbf{K}}$ are related by Eq. (74). For both channels, $\mathrm{Tr}(\rho_{\mathrm{m}}^R) \approx (Z(\mathbf{K}))^R$ when $R \to \infty$. This gives the $R \to \infty$ GPN duality.

The proof of Eq. (E9) parallels that of the CSS case. Consider HTE of $Z(\mathbf{K}, E)$ and LTE of $W(\mathbf{K}, E)$,

$$Z(\mathbf{K}, E) \propto \sum_{C \in (\mathcal{V}_s^*)^\perp} \tilde{\lambda}_x^{N_x(C)} \tilde{\lambda}_y^{N_y(C)} \tilde{\lambda}_z^{N_z(C)} (-1)^{C \cdot \Omega E}, \tag{E11}$$

$$W(\mathbf{K}, E) \propto \sum_{C \in \mathcal{V}_s} \lambda_x^{N_x(C)} \lambda_y^{N_y(C)} \lambda_z^{N_z(C)} (-1)^{C \cdot \Omega E}. \tag{E12}$$

In both cases, the proportionality constant is positive and independent of E. Thus, both lines are upper bounded by the $E = 0$ case $Z(\mathbf{K})$.

3. Connection to BPD duality

To make a connection to the CSS case, consider the following setup: $\lambda_x \neq 0, \lambda_y = \lambda_z = 0$. This describes pure bit-flip decoherence. The matching of Boltzmann weights [Eq. (E8)] for $R = 2$ gives

$$\frac{2\lambda_x}{1 + \lambda_x^2} = \left(\frac{1 + \tilde{\lambda}_x - \tilde{\lambda}_y - \tilde{\lambda}_z}{1 + \tilde{\lambda}_x + \tilde{\lambda}_y + \tilde{\lambda}_z} \right)^2$$

$$0 = \left(\frac{1 - \tilde{\lambda}_x + \tilde{\lambda}_y - \tilde{\lambda}_z}{1 + \tilde{\lambda}_x + \tilde{\lambda}_y + \tilde{\lambda}_z} \right)^2$$

$$0 = \left(\frac{1 - \tilde{\lambda}_x - \tilde{\lambda}_y + \tilde{\lambda}_z}{1 + \tilde{\lambda}_x + \tilde{\lambda}_y + \tilde{\lambda}_z} \right)^2. \tag{E13}$$

Solving the above equation, we get the error rates of the dual theory at $R = 2$,

$$\tilde{\lambda}_x = 1, \quad \tilde{\lambda}_y = \tilde{\lambda}_z, \quad \frac{2\lambda_x}{1 + \lambda_x^2} = \left(\frac{1 - \tilde{\lambda}_z}{1 + \tilde{\lambda}_z} \right)^2. \tag{E14}$$

The dual decoherence channel $\mathcal{E} = \otimes_\mu \mathcal{E}_\mu$ factorizes into independent bit-flip and phase-flip

errors,

$$\mathcal{E}_\mu(\rho_0) = \mathcal{N}_{z,\mu} \circ \mathcal{N}_{x,\mu}(\rho_0),$$

$$\mathcal{N}_{x,\mu}(\rho_0) = \tfrac{1}{2}\rho_0 + \tfrac{1}{2}X_\mu \rho_0 X_\mu, \tag{E15}$$

$$\mathcal{N}_{z,\mu}(\rho_0) = (1 - p'_z)\rho_0 + p'_z Z_\mu \rho_0 Z_\mu,$$

where p'_z is defined by $\tilde{\lambda}_z = \frac{p'_z}{1 - p'_z}$ (note $p'_z \neq \tilde{p}_z$). For CSS code, $\mathcal{V}_s = \mathcal{V}_x \oplus \mathcal{V}_z$ and $\mathrm{Tr}(\rho^R)$ factorizes under independent bit-flip and phase-flip errors [Eq. (12)],

$$\mathrm{Tr}(\rho^R) = \mathrm{Tr}(\rho_b^R) \times \mathrm{Tr}(\rho_p^R). \tag{E16}$$

In this case, the bit-flip part is a constant factor while the phase-flip error part depends on the parameter $\tilde{\lambda}_z$ determined by Eq. (E14). This relation is precisely the BPD duality for

CSS code at $R = 2$ [cf. Eq. (B8)]. The discussion for $R = \infty$ is analogous.

4. Self-dual surface

The self-dual surface of $R = \infty$ is the set of points on $p_x + p_y + p_z = \frac{1}{2}$ since any points on this surface satisfies

$$1 - 2p_y - 2p_z = p_x/p_0, \tag{E17}$$

and similarly upon cyclic permutations of p_x, p_y, p_z. In terms of K_x, K_y, K_z, one can rediscover the duality in the statistical model:

$$e^{-2(K_x+K_y)} + e^{-2(K_y+K_z)} + e^{-2(K_x+K_z)} = 1. \tag{E18}$$

To see that this defines the self-dual surface of the statistical model, one can apply Eq. (E18) to get

$$
\begin{aligned}
e^{-2(K_y+K_z)} &= \frac{1 + e^{-2(K_y+K_z)} - e^{-2(K_x+K_z)} - e^{-2(K_x+K_y)}}{1 + e^{-2(K_y+K_z)} + e^{-2(K_x+K_z)} + e^{-2(K_x+K_y)}} \\
&= \frac{e^{K_x+K_y+K_z} + e^{K_x-K_y-K_z} - e^{K_y-K_x-K_z} - e^{K_z-K_x-K_y}}{e^{K_x+K_y+K_z} + e^{K_x-K_y-K_z} + e^{K_y-K_x-K_z} + e^{K_z-K_x-K_y}} \\
&= \frac{(e^{K_x} - e^{-K_x})(e^{K_y} + e^{-K_y})(e^{K_z} + e^{-K_z}) + (e^{K_x} + e^{-K_x})(e^{K_y} - e^{-K_y})(e^{K_z} - e^{-K_z})}{(e^{K_x} + e^{-K_x})(e^{K_y} + e^{-K_y})(e^{K_z} + e^{-K_z}) + (e^{K_x} - e^{-K_x})(e^{K_y} - e^{-K_y})(e^{K_z} - e^{-K_z})} \\
&= \frac{\tanh K_x + \tanh K_y \tanh K_z}{1 + \tanh K_x \tanh K_y \tanh K_z}.
\end{aligned}
\tag{E19}
$$

The statistical model for $R = 2$ is equivalent to a single copy of the statistical model with renormalized K's, so we would expect a self-dual surface in that case as well. Comparing Eqs. (E6), (E7), and LTE of $Z_{\mathrm{SM}}(\mathbf{K})$, one can identify

$$(1 - 2p_y - 2p_z)^2 = e^{-2(K_y+K_z)}, \tag{E20}$$

and similarly for cyclic permutations of x, y, z. So the self-dual surface is determined by [cf. Eq. (E18)]

$$(1 - 2p_y - 2p_z)^2 + (1 - 2p_x - 2p_z)^2 + (1 - 2p_x - 2p_y)^2 = 1$$

$$\Leftrightarrow (1 - p_x - p_y - p_z)^2 + p_x^2 + p_y^2 + p_z^2 = \tfrac{1}{2}. \tag{E21}$$

APPENDIX F: PARENT HAMILTONIAN IN THE DOUBLED HILBERT SPACE

For a general stabilizer code $\mathcal{C}$, a stabilizer takes the general form

$$A_J[X, Z] = (\mathbf{i})^{\mathsf{a}_J \cdot \mathsf{b}_J} \prod_\mu (X_\mu)^{(\mathsf{a}_J)_\mu} \prod_\mu (Z_\mu)^{(\mathsf{b}_J)_\mu}, \tag{F1}$$

where a_J, b_J are $\mathbb{Z}_2$ vectors. Here, we choose to write the CSS code Hamiltonian as a sum of projectors so that the ground state has zero energy:

$$H_s = \sum_J \frac{\mathbb{1} - A_J}{2}. \tag{F2}$$

Notice that this Hamiltonian is related to the CSS code Hamiltonian in the main text by a factor of $\frac{1}{2}$ and some constant shift.

The Choi-Jamiołkowski isomorphism maps the system's density matrix into a state in the doubled Hilbert space. More

specifically, we can first choose the basis $\{|i\rangle\}$ of the original Hilbert space given by the eigenstates of all the Pauli-Z operators. The Choi-Jamiołkowski isomorphism is specified by the mapping $|i\rangle \langle j| \to |ij\rangle\rangle$, for all basis i, j. Therefore, if we have a density matrix represented as $\sum_{ij} \rho_{ij} |i\rangle \langle j|$, then the Choi-Jamiołkowski isomorphism maps it to its Choi representation $|\rho\rangle\rangle = \sum_{ij} \rho_{ij} |ij\rangle\rangle$.

One can multiply a density matrix with operators to modify it. It will, therefore, be crucial to spell out how to map these operations in the Choi representation. The most general operator action on the density matrix can be written as a sum of the following basic operation:

$$A\rho B = A_{ij}\rho_{jk}B_{kl}|i\rangle\langle l|. \tag{F3}$$

Therefore, in the Choi representation, the resulting action is given by

$$A_{ij}\rho_{jk}B_{kl}|il\rangle\rangle = A \otimes B^{\mathsf{T}}|\rho\rangle\rangle, \tag{F4}$$

where the transpose is done in the eigenbasis of Z's. For the pure-state density matrix ρ_0 of a logical state, it satisfies $H_s\rho_0 = 0$ and $\rho_0 H_s = 0$. In the Choi representation, this statement implies that $|\rho_0\rangle\rangle$ is the ground state of the following Hamiltonian:

$$H_0^D = H_s \otimes \mathbb{1} + \mathbb{1} \otimes H_s^{\mathsf{T}}. \tag{F5}$$

For the error-corrupted density matrix, its Choi representation in the doubled Hilbert space takes the form

$$|\rho_{\mathrm{m}}\rangle\rangle \propto \left(e^{\sum_\mu \tilde{K}_x X_\mu \otimes X_\mu - \tilde{K}_y Y_\mu \otimes Y_\mu + \tilde{K}_z Z_\mu \otimes Z_\mu}\right)|\rho_0\rangle\rangle \equiv \hat{\mathcal{E}}|\rho_0\rangle\rangle,$$

where we have defined the operator $\hat{\mathcal{E}} = e^{\sum_\mu \tilde{K}_x X_\mu \otimes X_\mu - \tilde{K}_y Y_\mu \otimes Y_\mu + \tilde{K}_z Z_\mu \otimes Z_\mu}$ acting on the doubled Hilbert space. One can construct a parent Hamiltonian for $|\rho_m\rangle\rangle$ by noticing that $|\rho_m\rangle\rangle$ is the ground state of the following frustration-free Hamiltonian as has been done in Ref. [20]

$$H^{D'} = \sum_J \left[\hat{\mathcal{E}}^{-1} \left(\frac{\mathbb{1} - A_J}{2} \otimes \mathbb{1} \right) \hat{\mathcal{E}} \right] \left[\hat{\mathcal{E}} \left(\frac{\mathbb{1} - A_J}{2} \otimes \mathbb{1} \right) \hat{\mathcal{E}}^{-1} \right] + \left[\hat{\mathcal{E}}^{-1} \left(\mathbb{1} \otimes \frac{\mathbb{1} - A_J^\top}{2} \right) \hat{\mathcal{E}} \right] \left[\hat{\mathcal{E}} \left(\mathbb{1} \otimes \frac{\mathbb{1} - A_J^\top}{2} \right) \hat{\mathcal{E}}^{-1} \right]. \quad (F6)$$

Note that every term in $H^{D'}$ is positive-semidefinite and annihilates the state $|\rho_m\rangle\rangle$. Hence, $|\rho_m\rangle\rangle$ must be a ground state of $H^{D'}$.

Recall that with a general stabilizer $A_J[X, Z]$, we associate two binary vectors $\mathsf{a}_{J,\mu}$ and $\mathsf{b}_{J,\mu}$. We can then calculate the individual terms as

$$\hat{\mathcal{E}} \left(\frac{\mathbb{1} - A_J}{2} \otimes \mathbb{1} \right) \hat{\mathcal{E}}^{-1} = \frac{1}{2} \mathbb{1} \otimes \mathbb{1} - \frac{1}{2} e^{2 \sum_\mu \tilde{K}_x \mathsf{b}_{J,\mu} X_\mu \otimes X_\mu - \tilde{K}_y (\mathsf{a}_{J,\mu} + \mathsf{b}_{J,\mu}) Y_\mu \otimes Y_\mu + \tilde{K}_z \mathsf{a}_{J,\mu} Z_\mu \otimes Z_\mu} A_J \otimes \mathbb{1}.$$

For later convenience, we introduce the following operator:

$$\hat{\mathcal{E}}_J = e^{\sum_\mu \tilde{K}_x \mathsf{b}_{J,\mu} X_\mu \otimes X_\mu - \tilde{K}_y (\mathsf{a}_{J,\mu} + \mathsf{b}_{J,\mu}) Y_\mu \otimes Y_\mu + \tilde{K}_z \mathsf{a}_{J,\mu} Z_\mu \otimes Z_\mu}. \quad (F7)$$

We remark that $(\mathsf{a}_{J,\mu} + \mathsf{b}_{J,\mu})$ should be understood as an addition in $\mathbb{Z}_2$. We thus have

$$\left[\hat{\mathcal{E}}^{-1} \left(\frac{\mathbb{1} - A_J}{2} \otimes \mathbb{1} \right) \hat{\mathcal{E}} \right] \left[\hat{\mathcal{E}} \left(\frac{\mathbb{1} - A_J}{2} \otimes \mathbb{1} \right) \hat{\mathcal{E}}^{-1} \right] = \left(\frac{1}{2} \mathbb{1} \otimes \mathbb{1} - \frac{1}{2} \hat{\mathcal{E}}_J^{-2} A_J \otimes \mathbb{1} \right) \left(\frac{1}{2} \mathbb{1} \otimes \mathbb{1} - \frac{1}{2} A_J \otimes \mathbb{1} \hat{\mathcal{E}}_J^{-2} \right)$$

$$= \frac{1}{4} \mathbb{1} \otimes \mathbb{1} + \frac{1}{4} \hat{\mathcal{E}}_J^{-4} - \frac{1}{4} (\hat{\mathcal{E}}_J^2 + \hat{\mathcal{E}}_J^{-2}) A_J \otimes \mathbb{1}.$$

Thus, we can write

$$H^{D'} = \sum_J \frac{1}{2} \hat{\mathcal{E}}_J^{-4} - \frac{1}{4} (\hat{\mathcal{E}}_J^2 + \hat{\mathcal{E}}_J^{-2}) (A_J \otimes \mathbb{1} + \mathbb{1} \otimes A_J^\top). \quad (F8)$$

A simpler parent Hamiltonian of $|\rho_m\rangle\rangle$ can be derived by noting that the double-Hilbert space operator $\hat{\mathcal{E}}$ is Hermitian and factorizes sitewise. So, we can write down a frustration-free Hamiltonian of the following form:

$$H^D = \sum_J \hat{\mathcal{E}}_J^{-1} \left(\frac{\mathbb{1} - A_J}{2} \otimes \mathbb{1} + \mathbb{1} \otimes \frac{\mathbb{1} - A_J^\top}{2} \right) \hat{\mathcal{E}}_J^{-1}. \quad (F9)$$

Note that, for each J, $\hat{\mathcal{E}}_J$ is so constructed that all the terms in the exponent anticommute with $A_J \otimes \mathbb{1}$ and also $\mathbb{1} \otimes A_J^\top$. One can then verify that

$$H^D \hat{\mathcal{E}} |\rho_0\rangle\rangle = \sum_J \hat{\mathcal{E}}_J^{-1} \hat{\mathcal{E}}_{\bar{J}} \left(\frac{\mathbb{1} - A_J}{2} \otimes \mathbb{1} + \mathbb{1} \otimes \frac{\mathbb{1} - A_J^\top}{2} \right) |\rho_0\rangle\rangle = 0, \quad (F10)$$

where $\hat{\mathcal{E}}_{\bar{J}}$ denotes all the terms in the original $\hat{\mathcal{E}}$ excluding the ones contained in $\hat{\mathcal{E}}_J$. $\hat{\mathcal{E}}_{\bar{J}}$ commute with $\mathbb{1} \otimes A_J^\top$ and $A_J \otimes \mathbb{1}$. Since each individual term of H^D is positive-semidefinite, $\hat{\mathcal{E}} |\rho_0\rangle\rangle = |\rho_m\rangle\rangle$ must be its ground state. The explicit form of H^D is obtained by moving $\hat{\mathcal{E}}_J$ across $\frac{\mathbb{1} - A_J}{2}$,

$$H^D = \sum_J \hat{\mathcal{E}}_J^{-2} - \frac{1}{2} A_J \otimes \mathbb{1} - \frac{1}{2} \mathbb{1} \otimes A_J^\top = H_s \otimes \mathbb{1} + \mathbb{1} \otimes H_s^\top + \sum_J \hat{\mathcal{E}}_J^{-2} + \text{const.} \quad (F11)$$

When $\tilde{K}_x, \tilde{K}_y, \tilde{K}_z$ are small, H^D to first order in $\tilde{K}$ is just H_0^D plus the interaction between two copies of the Hilbert space $\sum_J \sum_\mu \tilde{K}_x \mathsf{b}_{J,\mu} X_\mu \otimes X_\mu - \tilde{K}_y (\mathsf{a}_{J,\mu} + \mathsf{b}_{J,\mu}) Y_\mu \otimes Y_\mu + \tilde{K}_z \mathsf{a}_{J,\mu} Z_\mu \otimes Z_\mu$. This interaction competes with H_0^D. As we tune up the value of $\tilde{K}$'s, a possible scenario is that this competition leads to a quantum phase transition in H^D. Whether there is indeed a quantum phase transition in this model needs to be examined independently for different stabilizer codes.

---

[1] B. M. Terhal, Quantum error correction for quantum memories, Rev. Mod. Phys. **87**, 307 (2015).

[2] D. Gottesman, An introduction to quantum error correction and fault-tolerant quantum computation, in *Quantum information science and its contributions to mathematics*, Proceedings of Symposia in Applied Mathematics (2010), Vol. 68, pp. 13–58.

[3] A. Yu. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. **303**, 2 (2003).

[4] K. J. Satzinger, Y.-J Liu, A. Smith, C. Knapp, M. Newman, C. Jones, Z. Chen, C. Quintana, X. Mi, A. Dunsworth *et al.*, Realizing topologically ordered states on a quantum processor, Science **374**, 1237 (2021).

[5] R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, R. Babbush *et al.*, Suppressing quantum errors by scaling a surface code logical qubit, Nature (London) **614**, 676 (2023).

[6] T. I. Andersen, Y. D. Lensky, K. Kechedzhi, I. K. Drozdov, A. Bengtsson, S. Hong, A. Morvan, X. Mi, A. Opremcak, R. Acharya *et al.*, Non-Abelian braiding of graph vertices in a superconducting processor, Nature (London) **618**, 264 (2023).

[7] G. Semeghini, H. Levine, A. Keesling, S. Ebadi, T. T. Wang, D. Bluvstein, R. Verresen, H. Pichler, M. Kalinowski, R. Samajdar *et al.*, Probing topological spin liquids on a programmable quantum simulator, Science **374**, 1242 (2021).

[8] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, J. Math. Phys. **43**, 4452 (2002).

[9] H. G. Katzgraber, H. Bombin, and M. A. Martin-Delgado, Error threshold for color codes and random three-body Ising models, Phys. Rev. Lett. **103**, 090501 (2009).

[10] H. Song, J. Schönmeier-Kromer, K. Liu, O. Viyuela, L. Pollet, and M. A. Martin-Delgado, Optimal thresholds for fracton codes and random spin models with subsystem symmetry, Phys. Rev. Lett. **129**, 230502 (2022).

[11] C. Wang, J. Harrington, and J. Preskill, Confinement-Higgs transition in a disordered gauge theory and the accuracy threshold for quantum memory, Ann. Phys. **303**, 31 (2003).

[12] H. Bombin, R. S. Andrist, M. Ohzeki, H. G. Katzgraber, and M. A. Martin-Delgado, Strong resilience of topological codes to depolarization, Phys. Rev. X **2**, 021004 (2012).

[13] C. T. Chubb and S. T. Flammia, Statistical mechanical models for quantum codes with correlated noise, Annales de l'Institut Henri Poincaré D **8**, 269 (2021).

[14] R. Fan, Y. Bao, E. Altman, and A. Vishwanath, Diagnostics of mixed-state topological order and breakdown of quantum memory, PRX Quantum **5**, 020343 (2024).

[15] Y. Bao, R. Fan, A. Vishwanath, and E. Altman, Mixed-state topological order and the errorfield double formulation of decoherence-induced transitions, arXiv:2301.05687.

[16] J. Y. Lee, C.-M. Jian, and C. Xu, Quantum criticality under decoherence or weak measurement, PRX Quantum **4**, 030317 (2023).

[17] A. Yu. Kitaev, Quantum error correction with imperfect gates, in *Quantum Communication, Computing, and Measurement* (Springer, Berlin, 1997), pp. 181–188.

[18] S. Vijay, J. Haah, and L. Fu, Fracton topological order, generalized lattice gauge theory, and duality, Phys. Rev. B **94**, 235157 (2016).

[19] J. Haah, Local stabilizer codes in three dimensions without string logical operators, Phys. Rev. A **83**, 042330 (2011).

[20] J. Y. Lee, Y.-Z. You, and C. Xu, Symmetry protected topological phases under decoherence, arXiv:2210.16323.

[21] S. B. Bravyi and A. Yu. Kitaev, Quantum codes on a lattice with boundary, arXiv:quant-ph/9811052.

[22] H. Bombin and M. A. Martin-Delgado, Topological quantum distillation, Phys. Rev. Lett. **97**, 180501 (2006).

[23] H. Nishimori, Internal energy, specific heat and correlation function of the bond-random Ising model, Prog. Theor. Phys. **66**, 1169 (1981).

[24] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Clarendon Press, Oxford, UK, 2001).

[25] F. J. Wegner, Duality in generalized Ising models and phase transitions without local order parameters, J. Math. Phys. **12**, 2259 (1971).

[26] H. Nishimori, Duality in finite-dimensional spin glasses, J. Stat. Phys. **126**, 977 (2007).

[27] M.-D. Choi, Completely positive linear maps on complex matrices, Linear Algebra Appl. **10**, 285 (1975).

[28] A. Jamiołkowski, Linear transformations which preserve trace and positive semidefiniteness of operators, Rep. Math. Phys. **3**, 275 (1972).

[29] T. Ohno, G. Arakawa, I. Ichinose, and T. Matsui, Phase structure of the random-plaquette Z_2 gauge model: Accuracy threshold for a toric quantum memory, Nucl. Phys. B **697**, 462 (2004).

[30] Y. Ozeki and H. Nishimori, Phase diagram and critical exponents of the $\pm j$ Ising model in finite dimensions by Monte Carlo renormalization group, J. Phys. Soc. Jpn. **56**, 1568 (1987).

[31] N. Ito, Y. Ozeki, and H. Kitatani, Non-universal critical behavior in the ferromagnetic transition of the $\pm j$ Ising model, J. Phys. Soc. Jpn. **68**, 803 (1999).

[32] A. M. Ferrenberg and D. P. Landau, Critical behavior of the three-dimensional Ising model: A high-resolution Monte Carlo study, Phys. Rev. B **44**, 5081 (1991).

[33] K. Su, N. Myerson-Jain, C. Wang, C.-M. Jian, and C. Xu, Higher-form symmetries under weak measurement, Phys. Rev. Lett. **132**, 200402 (2024).

[34] P. Arnold and Y. Zhang, Monte Carlo study of very weakly first-order transitions in the three-dimensional Ashkin-Teller model, Nucl. Phys. B **501**, 803 (1997).

[35] D. A. Johnston and R. P. K. C. M. Ranasinghe, (four) dual plaquette 3D Ising models, Entropy **22**, 633 (2020).

[36] M. Mueller, W. Janke, and D. A. Johnston, Nonstandard finite-size scaling at first-order phase transitions, Phys. Rev. Lett. **112**, 200601 (2014).

[37] M. Kohmoto, M. den Nijs, and L. P. Kadanoff, Hamiltonian studies of the $d = 2$ Ashkin-Teller model, Phys. Rev. B **24**, 5229 (1981).

[38] G. Canossa, L. Pollet, M. A. Martin-Delgado, H. Song, and K. Liu, Exotic symmetry breaking properties of self-dual fracton spin models, Phys. Rev. Res. **6**, 013304 (2024).

[39] C. Chamon, Quantum glassiness in strongly correlated clean systems: An example of topological overprotection, Phys. Rev. Lett. **94**, 040402 (2005).

[40] Qubit stabilizer code, in *The Error Correction Zoo*, edited by Victor V. Albert and Philippe Faist (2022).

[41] F. Venn, J. Behrends, and B. Béri, Coherent-error threshold for surface codes from Majorana delocalization, Phys. Rev. Lett. **131**, 060603 (2023).

[42] D. Kribs, R. Laflamme, and D. Poulin, Unified and generalized approach to quantum error correction, Phys. Rev. Lett. **94**, 180501 (2005).

[43] D. W. Kribs, R. Laflamme, D. Poulin, and M. Lesosky, Operator quantum error correction, Quant. Inf. Comp. **6**, 382 (2006).

[44] M. B. Hastings and J. Haah, Dynamically generated logical qubits, Quantum **5**, 564 (2021).

[45] T. Rakovszky and V. Khemani, The physics of (good) LDPC codes I. Gauging and dualities, arXiv:2310.16032.

[46] Y. Zhao and D. E. Liu, Extracting error thresholds through the framework of approximate quantum error correction condition, arXiv:2312.16991.